

経路の回帰モデルによる旅行時間予測

Trajectory Regression for Travel-Time Prediction

井手 剛
Tsuyoshi Idé

IBM 東京基礎研究所
IBM Research - Tokyo
goodidea@jp.ibm.com, <http://www.trl.ibm.com/people/ide/>

加藤 整
Sei Kato

(同 上)
seikato@jp.ibm.com, <http://www.trl.ibm.com/people/seikato/>

keywords: trajectory mining, string kernel, Gaussian process regression

Summary

We propose a new method for predicting the travel-time along an arbitrary path between two locations on a map. Unlike traditional approaches, which focus only on particular links with heavy traffic, our method allows probabilistic prediction for arbitrary paths including links having no traffic sensors. We introduce two new ideas: to use string kernels for the similarity between paths, and to use Gaussian process regression for probabilistic prediction. We test our approach using traffic data generated by an agent-based traffic simulator.

1. はじめに

最近のセンシングおよび情報技術の進歩により、自動車や人間といった移動体を、広い領域にわたり追跡しその軌跡をデータとして蓄積することが可能になってきた。環境問題に絡んで最近重要性を増している高度道路交通システム (Intelligent Transportation System; ITS) は、まさにそのようなデータを提供する枠組みである。ちょうど Web ネットワーク上の情報流が Web マイニングという新しい研究分野を作り出したように、交通データも、その多様性と巨大さから新たな研究領域を生み出しつつある。

この論文では、ある所与の始点と終点の間の旅行時間予測 (または所要時間予測) という問題を考える。これは交通モデリングの最も基本的なタスクのひとつであり、現代的な ITS が現れるはるか以前、すでに 1970 年代には重要な研究課題として認識されていた [Ahrmed 79]。伝統的にはこの問題は、高速道路などのある特定のリンク (隣接する交差点の間の道路) に着目して、その地点で交通流の時系列モデリングを行うという形で扱われるのが通例であった。そもそも交通データでは、道路に設置したセンサーにより取得する以外の方法は長らく存在せず、いわば「路傍の視点」、すなわち、視点を地図上にひとつ固定してその場所で移動体の流れを観察する視点から現象を考えるのはいわば必然であった。

しかしいまや事情は変わった。GPS (Global Positioning Systems) など取得できる広域的なデータは、都市全体を俯瞰的に見た上で、任意の経路に対する何らかの予測

を行う手法により適しているように見える。例えば、渋滞回避のための迂回路を交通管制センターが各車両に指示するといった状況がそれであり、対象を自動車以外にも広げて想像してみると、例えば店舗内動線解析など、さらに多くの例があることに気づく。いずれも、従来研究が行っていたような、少数の特定リンクに対し精緻なモデルを立てるだけでは対応できない種類の、新しい問題であると考えざるを得ない。

新しい方向性を模索する上で、流体力学の定式化に 2 つの流派があるという事実は示唆的である。Euler 流のアプローチでは、空間の各点で流体の動態を記述する。一方 Lagrange 流の見方では、流体を粒子の集合と見て、それぞれの粒子を追跡することで運動の方程式を立てる [今井 70]。明らかに、路傍の視点は Euler 流の定式化に対応している。だとすれば、Lagrange 流の見方に対応するのは何か。

本論文が切り拓こうとするのはまさにその道である。我々は、ひとつの移動体に視点を置き、通過リンクを一覧する視点、すなわち「運転者の視点」に立脚して、旅行時間予測というタスクを考える。流体力学における 2 つのアプローチがそうであるように、路傍の視点と運転者の視点は、現象を見る仕方としては本質的には等価なものである。しかし暗黙に連続性を仮定する流体と異なり、交通流は一般には疎である。大都市圏であっても、よほどの幹線道路以外は、路傍の視点による時系列モデリングは、サンプル数の不足により困難であるし、そもそも、全リンクにおいて直接交通流をモデリングするのは、手間や時間を考えた時に不可能に近い。交通量情報の取得に

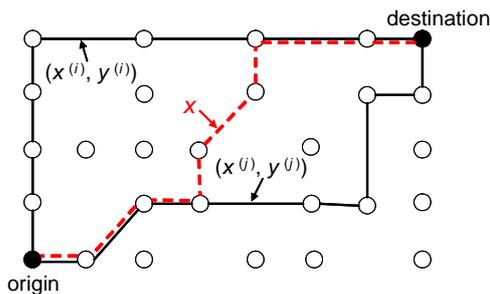


図 1 問題設定．過去の履歴が、経路と所要時間の組の集合として $\{(x^{(n)}, y^{(n)}) | n = 1, \dots, N\}$ のように与えられている時、点線で示されているような任意の経路 x に対して所要時間 y を予測する．

必要なインフラ自体が未整備という場合もよくある^{*1}．実用的な要請からは、運転者の視点に利があると言えよう．

我々の問題設定を改めて図 1 に示す．クエリとして与えられる経路 x は、一般に裏道なども含み、従来手法では旅行時間予測が難しい．また、可能な経路の個数はリンクの数に対し指数関数的であり、履歴の中に x と同一の経路を見出せるとは限らない．経路 x に対して実数値 y を関係付けるという意味では、これは回帰の問題と見なせるが、入力 x は普通のベクトルデータではないので、標準的な回帰公式は使えない．

そこで本論文では、文字列カーネルと正規過程回帰 (Gaussian process regression) を用いて旅行時間予測の問題を解くことを提案する．まず、経路を適当な文字列で表現し、経路同士の類似度を文字列カーネル [Leslie 02] で定義する．そして、ノンパラメトリックなカーネル回帰手法としての正規過程回帰を用いて、旅行時間を予測する．

やや抽象的に考えると、我々の問題は、移動体の経路、あるいはトラジェクトリと、旅行時間を結びつけるという問題である．トラジェクトリの解析はデータマイニングにおいて最近注目を集めている新しい研究分野であり [Vlachos 05, Lee 07, Kriegel 08]、GPS をはじめとしたセンシング技術の進歩にともない、今後応用面・理論面での研究の進展が期待されている．著者の知る限り、本研究はトラジェクトリマイニングの問題として旅行時間予測を扱った最初の仕事である．

2. 問題設定

本節では改めて旅行時間予測という問題を整理し、定式化における基本的前提について簡単に述べる．

2.1 定義

改めてリンクと経路の定義を与えよう．

【定義 1】(リンク) リンクとは隣接する交差点の間にある道路のことである．

【定義 2】(経路) 経路とは、任意の 2 つの連なったリンクがひとつの交差点を共有するようなリンクの系列である．

【定義 3】(旅行時間予測問題) 本論文での旅行時間予測タスクとは、訓練データ D から、任意の経路 x に対する旅行時間 y の確率分布 $p(y|x, D)$ を学習することである．

ここで、訓練データ D は過去の N 個の履歴情報からなり、

$$D \equiv \{(x^{(n)}, y^{(n)}) | n = 1, 2, \dots, N\} \quad (1)$$

と書かれる．ここで $x^{(n)}$ は、第 n 番目の履歴の経路であり、 $y^{(n)}$ はその経路に対する所要時間である． $y^{(n)}$ は、たとえば 312 秒というような実数となるが、 $x^{(n)}$ の方はリンクを表すシンボルの系列として与えられる．シンボルとしてリンク ID そのものを使った場合、たとえば

$$x^{(n)} = (25020201, 24021102, 222020101, 258020001, \dots)$$

のようなものである．リンク ID を含む道路ネットワークの情報は、国土地理院提供の電子地図から取得することができる．

2.2 前提とすること

以下、問題設定に関する基本的な注意を述べる．

交通データは一般に定常ではないので、 D はある時間区間を指定した上で収集されたと想定する．また、 x と、 D の中のすべての経路は、ある同一の始点と終点を通るものと仮定しておく．これは本論文の定式化の上での必須の要件というわけではないが、一般に、予測したい経路 x は、交通履歴 D と何らかの共通性を持っている必要があり、ここではその制約の一形態として、このような条件を想定するというのである．

先に述べたように、可能な経路の種類は指数個あるので、 x が D に含まれるとは限らないことに注意する．従って、「同じ経路を通った過去 100 台の平均」、のような単純な手法は一般には使えない．同様の理由で、「路傍の視点」による従来の定式化は、この問題を解くのに直接は使えない．交通履歴のないリンクに対する予測能力がないからである．

上記のように、本論文では、旅行時間の確率的予測を考える．言い換えると、予測そのもの（を期待値として与えること）に加えて、ひとつの予測に対するリスクをも出力したい．カーナビゲーションシステムを想定すればわかるように、これは実用的には非常に有用な情報である．本論文の定式化は、そのための首尾一貫した枠組みを提供する．ただし、公開された交通流データがほとん

*1 実際、名古屋市においては、幹線道路に限ってみても、VICS (vehicle information and communication system) ビーコンが設置されているのは道路長にしてわずか 22%のみである [Morikawa 07]

どない現時点では、予測のばらつきを実験とつぎ合わせて検証するのは困難であることを最初に述べておく。

3. 経路に対する回帰モデル

本節では、ある経路 x を入力とした回帰モデルの説明を行う。ポイントは、入力 x を明示的に何かの特徴ベクトルとして扱うのではなく、入力空間における類似度を介して回帰モデルを構築することである。これはカーネル回帰の枠組みによって実行できる。ここでは非線形関数の記述能力や結果の安定性、さらに確率的出力の得やすさの点から [Bishop 06]、正規過程回帰を採用する。

3.1 正規過程回帰

正規過程回帰における最初の仮定は、 \mathcal{D} における第 n 番目の旅行時間 $y^{(n)}$ の分布が、分散 σ^2 の観測ノイズを持つ正規分布により

$$p(y^{(n)}|f_n) = \mathcal{N}(y^{(n)} - \bar{y}|f_n, \sigma^2) \quad (2)$$

と表されるということである。ここで $\mathcal{N}(\cdot|f_n, \sigma^2)$ は平均 f_n 、分散 σ^2 の正規分布を表す。後で述べるように、 σ はデータから決められる上位パラメータ (hyperparameter) であるが、しばらく所与の定数と考えておく。 \bar{y} は \mathcal{D} における所要時間の総平均である。ノンパラメトリック回帰モデルの一般的特徴に従って、このモデルでは、各 $y^{(n)}$ に潜在変数 $f^{(n)}$ が結び付けられており、原理的には \mathcal{D} のいかなる関数関係でも再現することができる。しかしこのままでは過適合は必然であるから、正規過程回帰では、 $f^{(1)}, \dots, f^{(N)}$ に対して次の事前分布を置く。

$$p(\mathbf{f}_N) = \mathcal{N}(\mathbf{f}_N|\mathbf{0}, \mathbf{K}) \quad (3)$$

ここで $\mathbf{f}_N \equiv (f_1, \dots, f_N)^\top \in \mathbb{R}^N$ であり、共分散行列 \mathbf{K} の (i, j) 成分は経路 $x^{(i)}$ と $x^{(j)}$ の間のカーネル関数 $k(x^{(i)}, x^{(j)})$ として定義される (引数には入れていないが、 $p(\mathbf{f}_N)$ は $\{x^{(n)}\}$ に依存していることに注意)。

式 (2) および (3) という二つが正規過程回帰のモデルである。 y についての予測分布は、通常のベイズ推論の枠組みに従い、まず、潜在変数 f についての事後分布を求め、それを元に $p(y|f)$ から f を積分消去する。詳細は [Bishop 06] または [Idé 09] を参照されたい。これらの計算はすべて解析的に実行できて、予測分布は以下のように与えられる。

$$p(y|x, \mathcal{D}) = \mathcal{N}(y|m, s^2) \quad (4)$$

$$m = \bar{y} + \mathbf{k}^\top \mathbf{C}^{-1} \mathbf{y}_N \quad (5)$$

$$s^2 = \sigma^2 + k(x, x) - \mathbf{k}^\top \mathbf{C}^{-1} \mathbf{k} \quad (6)$$

ただし \mathbf{y}_N と \mathbf{k} は

$$\mathbf{y}_N = (y^{(1)} - \bar{y}, \dots, y^{(N)} - \bar{y})^\top \quad (7)$$

$$\mathbf{k} = (k(x^{(1)}, x), \dots, k(x^{(N)}, x))^\top \quad (8)$$

で定義され、 $\mathbf{C} \in \mathbb{R}^{N \times N}$ は

$$\mathbf{C} = \mathbf{K} + \sigma^2 \mathbf{I}_N \quad (9)$$

で定義される。ここで、 \mathbf{I}_N は N 次元の単位行列である。

3.2 文字列カーネル

二つの経路の類似度を考える際、最も素朴な方法は経路長に注目することである。我々の文脈ではこれは、長い経路には大きな旅行時間を期待する、ということの意味する。これは高速道路などではおそらく悪くないモデリングであると考えられるが、市街地での予測では、交差点での挙動が大きく旅行時間に影響するはずであり、第 0 近似としての意味しか持たないものと想像される。

このような見方を一般化して、まず各経路を何らかの文字列で表す。そのアルファベットとしては、先に与えた ID や、リンクの方向を表す東西南北などの文字を想定することができる。次に、その文字列の長さ p の連続する部分文字列を考え、それをを用いて、 $x^{(i)}$ と $x^{(j)}$ の間のカーネル関数を次のように定義する。

$$k_p(x^{(i)}, x^{(j)}) = \beta \sum_{\mathbf{u} \in \Sigma^p} N_{\mathbf{u}}(x^{(i)}) N_{\mathbf{u}}(x^{(j)}) \quad (10)$$

これは p -spectrum カーネルとして知られているものである [Leslie 02]。上に使った記号の定義は次のとおりである。

- Σ はリンクを表現するのに使われている文字の集合である。
- Σ^p は長さ p の相連なる部分文字列の集合である。
- $N_{\mathbf{u}}(x^{(i)})$ は、ある経路 (文字の系列) $x^{(i)}$ における部分系列 \mathbf{u} の出現回数である。

係数 β は事前分布の分散の大きさを制御する定数で、次節の通り、 σ 同様データから決められる上位パラメータとして扱われる。

3.3 上位パラメータ σ および β の計算

ベイズ推論の枠組みにおいては、上位パラメータ σ と β は周辺化尤度を最大化することで決められる。通常これは勾配法により行われるが、今のモデルでは比較的計算効率のよい固定点方程式を導ける。これを示そう。今、 σ の代わりに

$$\gamma \equiv \sigma^2 / \beta$$

を使えば、対数周辺化尤度は次のように書かれる。

$$\begin{aligned} \psi(\gamma, \beta) &\equiv \ln \int d\mathbf{f}_N p(\mathbf{f}_N) \prod_{n=1}^N p(y^{(n)}|f_n) \\ &= -\frac{1}{2} \ln \det(\mathbf{C}_1) - \frac{1}{2\beta} \mathbf{y}_N^\top \mathbf{C}_1^{-1} \mathbf{y}_N - \frac{N}{2} \ln \beta \end{aligned}$$

最後の等式で定数項を省いた。行列 \mathbf{C}_1 は次のように定義される。

$$\mathbf{C}_1 \equiv \mathbf{K}_1 + \gamma \mathbf{I}_N$$

ここで, K_1 は $\beta = 1$ の時のカーネル行列である. 行列の微分に関する標準的な公式を使えば [Bishop 06], 最適解の条件は次のようになる.

$$0 = \frac{\partial \psi}{\partial \gamma} = \frac{1}{2\beta} \mathbf{y}_N^T \mathbf{C}_1^{-2} \mathbf{y}_N - \frac{1}{2} \text{tr}(\mathbf{C}_1^{-1}) \quad (11)$$

$$0 = \frac{\partial \psi}{\partial \beta} = -\frac{N}{2\beta} + \frac{1}{2\beta^2} \mathbf{y}_N^T \mathbf{C}_1^{-1} \mathbf{y}_N \quad (12)$$

それぞれを交替的に繰り返し解くことで解が得られる. 後者については, ある γ の値について, 解が解析的に求められることに注意する. すなわち,

$$\beta = \frac{1}{N} \mathbf{y}_N^T \mathbf{C}_1^{-1} \mathbf{y}_N \quad (13)$$

数値計算的な詳細は別論文 [Idé 09] を参照されたい.

3.4 算法の要約

訓練時. σ と β を次のように求める.

- (1) 入力: カーネル行列 K , 旅行時間のベクトル \mathbf{y}_N , σ と β についての初期値.
- (2) 手順: 式 (11) および (12) を交替的に収束するまで解く.
- (3) 出力: ψ を最大化する σ^2 と β .

予測時. あらかじめ C の Cholesky 因子 L とその逆 L^{-1} , さらに $\mathbf{h} \equiv L^{-1} \mathbf{y}_N$ を計算しておく.

- (1) 入力: 経路 x (と事前計算された L^{-1} および \mathbf{h}).
- (2) 手順:
 - $\mathbf{l} \equiv L^{-1} \mathbf{k}$ を計算する.
 - 予測平均 $m = \bar{y} + \mathbf{h}^T \mathbf{l}$ を計算する.
 - 予測分散 $s^2 = \sigma^2 + k(x, x) - \mathbf{l}^T \mathbf{l}$ を計算する.
- (3) 出力: 予測平均 m と予測分散 s^2 .

4. 実験

本節では実際の地図情報に基づく交通データに基づいて旅行時間予測の実験を行う.

4.1 実験の設定

データ D は, 実際の京都市街の電子地図の上で, エージェントベースのトラフィックシミュレータである IBM Mega Traffic Simulator *2 により生成した. 図 2 にシミュレータのスクリーンショットを示す. 画面で薄く点で表されているのが生成されたエージェント (自動車) である. その地図の左上が経路の始点, 右下が終点である.

各エージェントはそれぞれに対して事前に定義された経路をたどるが, 各リンクの法定速度に加えて, 他のエージェントとの相互距離により速度を変化させるようプログラムされているので, 得られる旅行時間は試行ごとに大きくばらつく. ここではまず, ある N_0 個の経路の候補

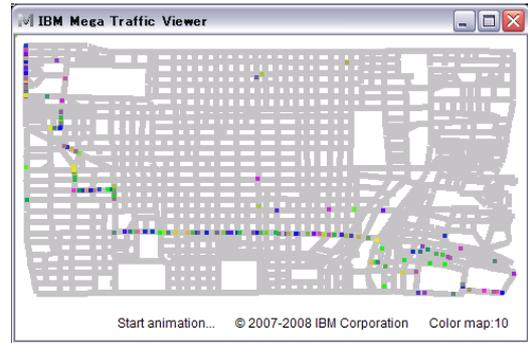


図 2 IBM Mega Traffic Simulator と, その画面上に表示された京都市街地図.

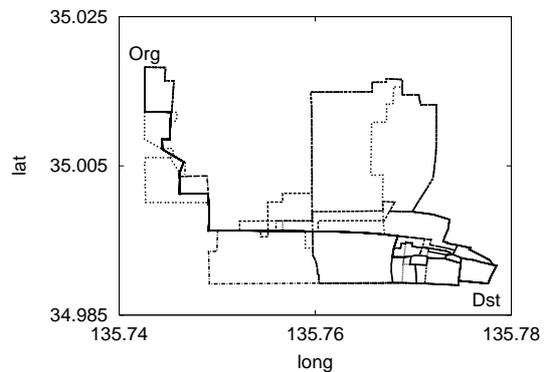


図 3 最短所要時間経路のいくつか. 横軸と縦軸はそれぞれ地図上の経度と緯度を表す.

を, 理性的な運転者の行動を模擬する目的で k 最短経路アルゴリズム [Yen 71] を用いて生成し (図 3), 各エージェントはその中から経路をランダムに選択するものとした. 「最短」の定義は, 法定速度に基づいて単独で走行した時に, 所要時間が最小になるということである. エージェントの発生間隔は, 平均 0.1 秒の指数分布に従うものと仮定し, 信号待ちを模擬するために, 交差点においては, τ 秒の待ち時間を与える. 以下の実験では, $N_0 = 132$ とし, 訓練用に $N = 100$ 個をランダムに選択し, 残りをテスト用とする*3.

4.2 比較対象

上に説明したデータに対し, 3 つのカーネル関数を比較する. (1) 文字列集合 Σ としてリンクの ID そのものを採用した p -spectrum カーネル, (2) Σ としてリンクの方角 (東西南北) を採用した p -spectrum カーネル, (3) 2 つの経路の囲む面積を非類似度に使ったカーネル, である. それぞれ, ID カーネル, 方向カーネル, 面積カーネルと呼ぶ. 我々は過去の交通履歴のない未知リンクを含みうる任意経路に対する予測能力を議論しているので, 従来の「路傍の視点」に基づく手法との比較は行えない

*2 <http://www.ibm.com/jp/press/pressroom/kaiken/20080610a.pdf>

*3 データの一部は次の URL で入手できる: http://www.trl.ibm.com/projects/socsim/project_e.htm

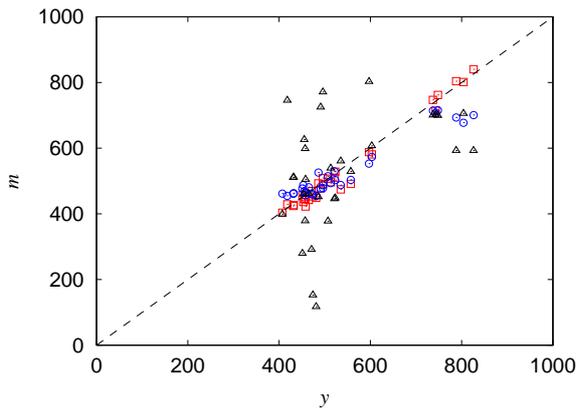


図4 予測旅行時間 (m) と実測値 (y) との比較. ID カーネルを \square , 方向カーネルを \circ , 面積カーネルを \triangle で表す.

ことに注意する.

面積カーネルは, 2 つの経路 $x^{(i)}$ と $x^{(j)}$ の囲む面積を $S(x^{(i)}, x^{(j)})$ とした時,

$$k^{\text{area}}(x^{(i)}, x^{(j)}) \equiv \beta e^{-S(x^{(i)}, x^{(j)})}$$

のように定義される. $S(x^{(i)}, x^{(j)})$ は地図上の 2 つの経路の ℓ_1 距離の拡張とみなすことができる. この意味で面積カーネルは, 動的時間伸縮法など 2 つの経路の間の非類似度を幾何学的方法で定義する方法の対応物とみなすことができる.

4.3 実験結果

ここではまず, 異なるカーネル同士の予測性能を比較し, 次いで ID カーネルの文字列長 p に対する予測精度の依存性を論じる.

異なるカーネルの比較. 図4に異なるカーネル同士の比較結果を示す. 待ち時間パラメータは $\tau = 10$ とし, ID カーネルと方向カーネルについては部分文字列長は $p = 2$ とした. 縦軸が予測値, 横軸が実測値であり, 点線が完全一致を示す. 図から, ID および文字列カーネルについては予測と実測の対応が比較的良好である. 他方, 面積カーネルの結果は顕著に悪い. これは, 地図上で経路同士の「形」を直接比べるような手法が役に立たないことを意味する. たとえば高速道路と一般道が併走している状況を想像すれば直感的に納得できる結果であろう. トラジェクトリマイニングのこれまでの研究では, 何らかの意味で実空間の「形」の比較に基づいて議論がなされることが多かった. しかしこの場合は, 実空間ではなく, p -spectrum カーネルが定めるヒルベルト空間において類似性を考えねばならないということであり, 興味深い結果と言える.

表1に上記の結果を要約してある. r は, テストデータにわたって, 旅行時間の実測値と式(5)の予測平均との相関係数を計算したものである. ID カーネルが相関係数にして 0.98 と, 最もよい値を示していることがわかる.

表1 異なるカーネル同士の r と平均 s^2 値の比較 ($\tau = 10$).

| | ID | direction | area |
|--------------|--------------|-----------|-------|
| r | 0.980 | 0.933 | 0.059 |
| $\sqrt{s^2}$ | 4.5 | 10.0 | 10.3 |

方向カーネルがそれに続き, 面積カーネルの r は極めて低い.

表にはテストデータにわたる予測分散(式(6))の平均の平方根の値も示してある. 表からわかるとおり, 方向カーネルと面積カーネルの予測分散はほぼ同じであるが, ID カーネルの予測分散は小さい. 言い換えると, ID カーネルを使うと, リスクが小さい予測ができるということである.

以上の結果から, 今回の実験では ID カーネルがもっとも好ましい結果を示した. これはそもそもの情報量の相違を考えるとある程度理解できる. ID カーネルはリンクを一意的に指し示すシンボルをカーネル計算に使っており, この意味で最も豊富な情報を持つ. 一方, 方向カーネルはリンクの絶対的な位置の情報を捨象して方向のみを見ているものの, 今回のデータでは, 始点と終点が共通の経路のみを含んでいることと, 京都市街地図が碁盤の目状であるという 2 つの理由により, 実際には, 東西南北の系列を与えても経路はほとんど一意に決まってしまう. これが ID カーネルと似た r 値を示した主な理由であろう. 一般に交通データにおいては外れ値やノイズとみなされるデータが多く含まれ, 従って, 方向カーネルのような抽象化がノイズ削減に積極的な効果をもたらすことを期待できるが, 今回の実験ではそのような効果は副次的なものであった. 最後に, 経路の絶対的な情報を完全に捨てて, 形の類似のみを見る面積カーネルでは, うまく旅行時間の予測を行うことができなかった.

異なる文字列長の比較. ここでは先の実験で最も良い予測性能を示した ID カーネルを対象に, いくつかの p の値に対して予測平均 $m(x)$ を計算し, 実測値との相関係数 r を求めた. 結果を図5に示す. 予測精度は τ の値によって異なるが, おおむね $p = 2$ で最も精度がよいことがわかる.

データを詳しく見ると, たとえば $\tau = 10$ 秒の時, 交差点での待ち時間は, 総旅行時間のおよそ 3 割ほどを平均して占めることがわかる. すなわち, 総旅行時間には個々のリンク単独の影響がまず支配的で, 次いで隣接する複数のリンクの関係が効いてくる. つまり右左折や直進など, 交差点での振る舞いの相違が結果に効いてくるということである. このことを考えると, $p = 2$ で最も高い性能を示すというのは理解できる. 交差点での走行パターンを考慮できる最も短い文字列長だからである.

理論的には, 文字列長 p の最適化は, カーネルパラメータを最適化もしくは学習する問題であり, 興味深い研究課題である. 今回はテストデータにわたる予測平均と

実測値との相関係数を用いて最適な p 値を同定したが、ベイズ理論の観点からより首尾一貫しているのは、 β や σ と同じく、周辺化尤度の最大化によりこれを求めることである。これについては別途報告することにした。

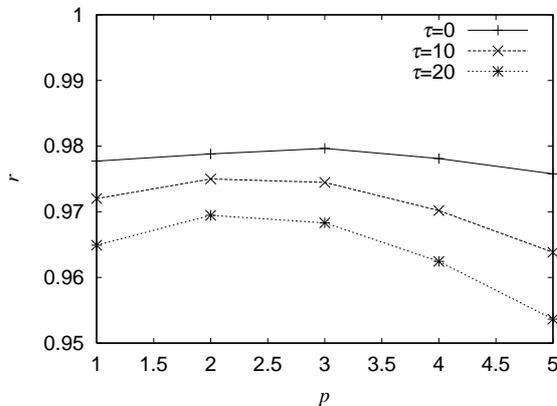


図5 文字列長 p の関数として表した r 値 (横軸の 1.5 などの端数には意味はない)。実線が $\tau = 0$ 、破線が $\tau = 10$ 、点線が $\tau = 20$ を示す。

5. ま と め

任意経路に対する旅行時間の予測という問題を設定し、トラジェクトリに対する回帰の問題として定式化した。提案した手法は未知リンクを含む経路についても旅行時間を予測できるという、従来技術にはなかった特徴を持つ。京都市街地図上でのシミュレーションデータを用いて、文字列カーネルに基づく正規回帰という我々の手法が十分な予測能力を持つことを検証した。

今後の課題としては、まず、実データでの検証とシミュレーションの精緻化が挙げられる。我々のシミュレータは、自由流相と渋滞相との間のメタ安定状態 [Nishinari 02] を再現する能力を持つという意味で現実の交通流の特徴を再現できるが、今回の実験の設定がどの程度現実的かは必ずしも明らかでない。

理論面でも、データのスケラビリティへの対応や、カーネル最適化手法の検討など興味深い課題がいくつか残されている。また、今回の実験では、ID カーネルが最も良い予測性能を示した。この結果は、経路の回帰という我々の問題の定式化を再考する上で示唆的であると思われる。これらの課題については、別途詳しく議論する予定である。

謝 辞

この研究の一部は、総務省の地球温暖化対策 ICT イノベーション推進事業 (PREDICT) の助成により行われた。謝意を表する。

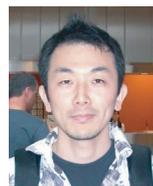
◇ 参 考 文 献 ◇

- [Ahrmed 79] Ahrmed, M. S. and Cook, A. R.: Analysis of freeway traffic time-series data by using Box-Jenkins techniques, *Transportation Research Record*, Vol. 722, pp. 1–9 (1979)
- [Bishop 06] Bishop, C. M.: *Pattern Recognition and Machine Learning*, Springer-Verlag (2006)
- [Idé 09] Idé, T. and Kato, S.: Travel-Time Prediction using Gaussian Process Regression: A Trajectory-Based Approach, in *Proc. SIAM Intl. Conf. Data Mining* (2009)
- [Kriegel 08] Kriegel, H.-P., Renz, M., Schubert, M., and Zuefle, A.: Statistical Density Prediction in Traffic Networks., in *Proc. SIAM Intl. Conf. Data Mining*, pp. 692–703 (2008)
- [Lee 07] Lee, J., Han, J., and Whang, K.-Y.: Trajectory Clustering: A Partition-and-Group Framework, in *Proc. 2007 ACM SIGMOD Intl. Conf. Management of Data*, pp. 593–604 (2007)
- [Leslie 02] Leslie, C., Eskin, E., and Noble, W. S.: The spectrum kernel: A string kernel for SVM protein classification, in *Proc. the Pacific Symposium on Biocomputing*, pp. 564–575 (2002)
- [Morikawa 07] Morikawa, T., Yamamoto, T., Miwa, T., and Wang, L.: Development and Performance Evaluation of Dynamic Route Guidance System PRONAVI, *Journal of the Japan Society of Traffic Engineers*, Vol. 42, No. 3, pp. 65–75 (2007)
- [Nishinari 02] Nishinari, K., Fukui, M., and Schadschneider, A.: A stochastic cellular automaton model for traffic flow with multiple metastable states, *Journal of Physics, A*, Vol. 71, pp. 2339–2347 (2002)
- [Vlachos 05] Vlachos, M.: Elastic Translation Invariant Matching of Trajectories, *Machine Learning Journal*, Vol. 58, No. 2-3, pp. 301–334 (2005)
- [Yen 71] Yen, J. Y.: Finding the K Shortest Loopless Paths in a Network, *Management Science*, Vol. 17, No. 11, pp. 712–716 (1971)
- [今井 70] 今井功: 流体力学, 岩波書店 (1970)

〔担当委員: 小野田 崇〕

2009 年 9 月 7 日 受理

著 者 紹 介



井手 剛(正会員)

国立苫小牧工業高等専門学校機械工学科、東北大学工学部機械工学科を経て、2000 年東京大学大学院理学系研究科・物理学専攻博士課程修了。同年 IBM 東京基礎研究所入所。液晶工学、データマイニングの研究に従事。現在、同研究所アドバイザー・リサーチャー。人工知能学会、電子情報通信学会、日本物理学会、ACM SIGKDD 各会員。



加藤 整(正会員)

東京大学工学部航空宇宙工学科、同大学院工学系研究科航空宇宙工学専攻を経て、2002 年同大学院数理科学研究科博士課程修了。同年 IBM 東京基礎研究所入所。リスク解析、マルチエージェントシステムの研究に従事。現在、同研究所主任研究員。情報処理学会、日本シミュレーション学会各会員。