

旅行時間予測値の 性能に関する統計分析

坂井 勝哉¹・Chong WEI²・日下部 貴彦³・朝倉 康夫⁴

¹学生会員 神戸大学 大学院工学研究科市民工学専攻 (〒657-8501 兵庫県神戸市灘区六甲台町1-1)
E-mail:katsuya-sakai@stu.kobe-u.ac.jp

²非会員 東京工業大学研究員 大学院理工学研究科土木工学専攻 (〒152-8552 東京都目黒区大岡山2-12-1-M1-20)
E-mail:c.wei@plan.cv.titech.ac.jp

³正会員 東京工業大学助教 大学院理工学研究科土木工学専攻 (〒152-8552 東京都目黒区大岡山2-12-1-M1-20)
E-mail:t.kusakabe@plan.cv.titech.co.jp

⁴正会員 東京工業大学教授 大学院理工学研究科土木工学専攻 (〒152-8552 東京都目黒区大岡山2-12-1-M1-20)
E-mail:asakura@plan.cv.titech.co.jp

本研究では、先行研究の統計モデルを用いた旅行時間推定モデルを拡張することにより、旅行時間予測値を区間推定するモデルを構築した。さらに、先行研究の統計モデルでは、2つの変数をパラメータとしていたが、これらのパラメータについて統計的な独立性を仮定した場合のモデルを構築し、そのモデルの挙動を調べた。これらのモデルを用いた分析を行い、構築したモデルを用いることで、旅行時間予測に用いる統計データの数が少ない場合の予測値の性能の低下などを確認できた。また、独立性を仮定したモデルを用いることで、独立性を仮定していないモデルの欠点である、過去のデータの少ない交通状態での予測も行えることが確認できた。

Key Words : *Travel time prediction, Reliability on prediction, Bayesian probability, MCMC*

1. 序論

所要時間情報は、利用者のニーズが高い交通情報の一つであり、様々な旅行時間予測モデルが提案されている。例えば、舟橋ら¹⁾は、VICSデータから現在の状況と類似した過去のデータを予測値として用いる手法を構築している。宗像ら²⁾は、リアルタイムの観測値を入力値として交通流シミュレーションを用いた予測方法を提案している。このように、旅行時間の予測モデルには、主に過去のデータに基づいた方法と交通流をモデル化したシミュレーションに基づく方法が提案されている。

本研究の先行研究であるWeiら³⁾の研究では、交通流の性質の一つであるボトルネックの性質を考慮した上で、過去の検知器によるデータを統計的に用いた旅行時間の予測手法を提案している。この手法は、ある時刻の検知器による速度データから算出される区間毎の旅行時間の同時和である瞬間旅行時間(ITT: Instantaneous Travel Time)と旅行時間予測対象区間の上流部から渋滞末尾までの距

離を入力値として、その時刻に流入した利用者が体験する旅行時間である実旅行時間(RTT: Real Travel Time)を予測するものである。この手法では、交通流の性質を考慮していることに加え、車両検知器によるデータから得られた過去の交通流データのみから予測モデルのパラメータを得ることができるため、車両検知器が設置されている区間では、比較的容易に実装できるという特徴がある。

Weiらの手法は、旅行時間の予測値の推定を点推定で行っており、入力値が与えられれば、唯一の予測値が出力される手法となっているが、Weiらのモデルを拡張すると、一定の確率で観測されることが期待される予測値の幅を推定する手法に拡張することを比較的容易にできると考えられている。このような予測値の幅の推定、すなわち予測値の区間推定が可能になることで、その予測結果の確からしさ、すなわち予測結果の性能を予測結果の値とともに道路利用者に提供できるようになると期待できる。このことは、例えば、区間推定で予測した旅行時間の平均値が80分の場合で、5~90%の予測区間の幅

が40分の場合と5分の場合があれば、道路利用者は予測区間の幅が5分の場合には旅行時間の平均値に近い旅行時間を想定した行動を実施できるが、予測区間の幅が40分の場合には予測結果を信頼できないことを考慮した行動をとることができるようになることを示している。また、予測区間の幅は、旅行時間情報の提供者が、どの予測方法を用いて情報提供を行うかを定める指標になると考えられる。例えば、予測区間の幅が大きい予測結果がでた場合には、より予測区間の幅が小さくなる予測モデルに切り替えて情報提供を行うことが可能になるだろう。そこで、本研究の一つ目の目的は、旅行時間の予測時に、その予測結果の性能を同時に推定できるモデルを作成するために、Weiらのモデルを、旅行時間予測値を点推定するモデルから区間推定するモデルに拡張することである。

Weiらのモデルは、交通流の性質を利用したモデルであったため、その拡張したモデルで、予測結果の性能が高くない結果が得られた場合、交通流の性質以外の要因が旅行時間の変動に働いている可能性が考えられる。このような場合には、旅行時間の変動に影響を与える変数を入力値に加えることになるだろう。例えば、Weiらのモデルの入力値に加えて、時間帯を変数として組み入れることで、時間帯毎の交通流の特徴の変化を反映できるモデルを作成できると考えられる。しかし、変数が多くなると、Weiらのモデルをそのまま拡張した場合には、入力値となる変数の値の組み合わせが膨大な数になり、これらの組み合わせすべてが過去の観測データに十分に含まれているとは限らない状況になって、統計的な方法での予測が難しくなると考えられる。この問題を解決する方法として、変数間の独立性を仮定することが考えられる。しかし、変数の独立性を仮定した場合の予測モデルの挙動は独立性を仮定しない場合と異なることが想定され、適切な予測値が推定できなくなる可能性もある。したがって、独立性を仮定することによる予測モデルへの影響を知ることが必要である。したがって、2つ目の目的は、Weiらのモデルの入力値の変数について、互いに独立であると仮定したモデルを構築するとともに、独立性を仮定しないモデルとの比較を行いモデルの挙動を確かめることである。

2. 旅行時間予測モデル

第2章では、Weiらの旅行時間モデルを拡張し、旅行時間の予測値を区間推定するモデルと、さらにそのモデルを拡張して、入力値の変数に関して統計的な独立性を仮定したモデルを構築する。第1節ではWeiらのモデルについて説明を行う。第2節では、区間推定をするモデル

を構築し、第3節で入力値の変数について統計的独立性を仮定したモデルに拡張する。第4節では、本章で構築したモデルの適用範囲について述べる。

(1) Weiらのモデル

Weiらは、予測の対象となる1区間に1つのボトルネックが存在する区間において、「ボトルネック容量が一定である」、「渋滞長の延伸（収縮）の速度が一定である」との仮定に基づいたモデルである。このモデルは、

$$RTT(t) = F(C(t), x(t)) \cdot ITT(t) \quad (1)$$

ただし、

$ITT(t)$: 時刻 t の瞬間旅行時間

$RTT(t)$: 時刻 t に旅行時間の予測対象区間に流入する車両の実旅行時間

τ : 車両検知器の観測周期

$$C(t) : \text{PIITの变化率であり } C(t) = \frac{ITT(t)}{ITT(t-\tau)}$$

$x(t)$: 旅行時間の予測対象区間の最上流部から渋滞末尾までの距離

で表すことができる。 $F(C(t), x(t))$ は、 $C(t)$ と $x(t)$ の値の組み合わせ毎に異なる値をもつ離散的なパラメータである。 $F(C(t), x(t))$ の値は、過去の検知器データから時刻毎に算出された $ITT(t)$ 、 $RTT(t)$ 、 $C(t)$ 、 $x(t)$ を、

$C(t)$ と $x(t)$ の値の組み合わせ毎に分割し、分割されたデータセット毎に式(1)に線形回帰を行うことで求めることができる。旅行時間の予測値である $RTT(t)$ を推定する際には、予測を行う時刻の検知器による観測値から、 $ITT(t)$ 、 $C(t)$ 、 $x(t)$ を求め、その $C(t)$ と $x(t)$ の値の組み合わせに対応する $F(C(t), x(t))$ を用いて式(1)より算出する。

(2) 区間推定を行うモデルへの拡張

第1節でのWeiらのモデルは、旅行時間の予測値について点推定を行うモデルであり、時刻 t の瞬間旅行時間 $ITT(t)$ 、PIITの变化率 $C(t)$ 、上流部から渋滞末尾までの

距離 $x(t)$ を入力すると唯一の旅行時間予測値が求まるものである。ここでは、旅行時間の予測結果とともに、その結果の性能を推定するために、旅行時間を点推定で推定するのではなく、区間推定を行うモデルを構築する。区間推定を行えば、例えば、推定された区間の10%タイル値と90%タイル値を用いて80%の確率で、実際に観測される旅行時間が入ると想定される旅行時間の幅を求めることができる。

区間推定を行うために、 $F(C(t), x(t))$ を確率変数 $F_{c,x}$ とし、 $F_{c,x}$ が $C(t)$ と $x(t)$ の値の組み合わせ毎に異なる確率密度関数 $p(F_{c,x}|C(t), x(t))$ に従うとする。本研究では、 $F_{c,x}$ が正規分布に従うし、その正規分布は $C(t)$ と $x(t)$ の値の組み合わせ毎に異なる平均と分散をもつものとする。

$F_{c,x}$ が従う正規分布の平均と分散は、過去の検知器データから時刻毎に算出された $ITT(t)$ 、 $RTT(t)$ 、 $C(t)$ 、 $x(t)$ を、 $C(t)$ と $x(t)$ の値の組み合わせ毎に分割し、データセット毎の平均値と分散から求めることができる。

区間推定を行うモデルについての予測モデルとしての性能には、2つのものが考えられる。一つ目は、モデルによって得られた個々の旅行時間の予測値そのものに関する性能であり、もう一つは、予測モデルとしての性能である。

区間推定で、予測結果が実際に観測された実旅行時間に対して的中したかの判断は、推定を行ったのと同じ時間帯に実際に観測された実旅行時間が、予測モデルによる区間推定結果を用いて一定の確率で観測されると想定される旅行時間の幅の中に入るかどうかを確認することによって判断できる。さらに、予測値に関する性能は、推定区間によって得られた一定の確率で観測されると想定される旅行時間の幅の広さによって示すことができる。この幅が広いということは、大小様々な旅行時間が含まれる予測結果を出しているということになり、実際に観測される旅行時間がそれらのどれになるかわからないということである。例えば、この幅が15分~120分であった場合には、実際に観測される実旅行時間がその幅の中のどの値をとっても予測が的中したことになる。したがって、このような区間が広い予測値は性能が良くないと判断できる。一方で、予測された区間が狭い場合には、

旅行時間情報の利用者はより確実に旅行時間を把握できることから、性能が良い予測値と判断できる。また、性能が良い予測値を多く出力できるモデルは性能が良いと判断できる。したがって、予測値の性能を用いることで予測モデルの性能を評価できると考えられる。また、実際に観測された実旅行時間が推定された区間内によく入るモデルは、予測モデルとしての性能が高いといえる。このことから、予測値の性能を用いることで予測モデルの性能が評価できることに加えて、推定区間への的中率を判断することでもモデルの性能を評価できる。つまり、これらの性能の評価を組み合わせると、性能のよいモデルとは、予測結果の推定区間が狭くかつ実際に観測された実旅行時間が必ず的中するモデルであるといえる。

(3) 変数間の独立性を仮定したモデルへの拡張

第2節で示したモデルでは、Weiらのモデルで交通流の性質を示す変数として導入されていた $C(t)$ と $x(t)$ の値の組み合わせ毎に $F_{c,x}$ が従う分布を推定していた。

しかし、Weiらのモデルで想定している交通流の性質のみで、旅行時間の変動を説明できない場合もあり得る。このようなケースは、例えば、日照や天候などによってボトルネック容量が変動する場合などが相当するだろう。このような場合には、その変動を説明できる変数を導入することで推定の精度を向上させることができると考えられる。一方で、変数の数が増えると、変数の値の組み合わせの数も大きく増えることから、第2節での $F_{c,x}$ に相当する確率変数が従う分布を推定する際に用いる過去のデータにその値の組み合わせが生じていない場合や生じていてもごく少数となっていることが多分に生じると考えられる。このような場合には、確率変数が従う分布がうまく推定できず、予測モデルの性能が悪くなることが懸念される。このような、問題を防ぐ手立てとして、

$F_{c,x}$ に相当する確率変数が従う分布が、モデルへの入力値となる変数について互いに独立であるようにモデルを構築することが考えられる。つまり、第2節では、 $C(t)$ と $x(t)$ の組み合わせ毎に、 $F_{c,x}$ の分布を定義していたが、 $C(t)$ と $x(t)$ について独立性を仮定する場合には、 $C(t)$ と $x(t)$ それぞれについて、確率変数 F_c 、 F_x を定義する。このとき、 $F_{c,x}$ が従う確率密度関数

$p_i(F_{c,x}|C(t), x(t))$ は、 F_c が従う確率密度関数 $p(F_c|C(t))$ と F_x が従う確率密度関数 $p(F_x|x(t))$ を用いて、

$$p_i(F_{c,x}|C(t), x(t)) = p(F_c|C(t))p(F_x|x(t)) \quad (2)$$

として表せる。なお、 F_c 、 F_x はそれぞれ正規分布に従うと仮定する。これらの分布のパラメータである平均値と分散の推定の際には、過去の検知器データを $C(t)$ の値毎に分割したデータを用いて F_c の分布のパラメータを求め、 $x(t)$ の値毎に分割したデータを用いて F_x のパラメータを求める。

(4) モデルの適用範囲

第1～3節でのモデルは、渋滞が発生していることを前提として構築されたモデルである。また、もし、渋滞が発生していない場合には、速度の変動が小さいため「 $RTT(t) = ITT(t)$ 」となることが期待され、本研究のような予測の必要性は少ないと考えられる。一方で、渋滞している場合には、速度の変動が渋滞の延伸などに多じて大きくなるため「 $RTT(t) \neq ITT(t)$ 」であり、旅行時間予測の必要性も大きく、本研究での旅行時間予測の対象となる。渋滞の発生を前提としているため、予測区間にボトルネックが含まれている必要があるが、本研究で構築したモデルは、第1節でのWeiらのモデルを拡張したものであるため、そのモデルと同様に、予測の対象となる1区間に1つのボトルネックが存在する区間に適用できるモデルとなっている。

3. 実データを用いたパラメータ推定と推定結果によるモデルの特性の分析

この章では、第2章で提案したモデルのパラメータ推定を行う。第1節では、使用したデータや操作について述べる。第2節ではパラメータの推定方法、第3節ではパラメータの推定結果を示す。最後に第4節で $C(t)$ と $x(t)$ の独立性を仮定しないモデルと仮定するモデルの結果を比較する。

(1) データ

対象区間は阪神高速3号神戸線上り尼崎料金所から阿波座ジャンクションまでの約8kmとする。この区間にはおよそ500m間隔で18箇所に検知器が設置されており、5分間隔で、交通量・速度・占有率のデータが取得されている。したがって、式(1)での τ は、 $\tau = 5$ である。この区間の最下流部には合流によるボトルネックがあり、1区間に1つのボトルネックが存在するという第2章4節で述べた適用範囲の条件を満たしている。

2003年3月から2010年6月までに検知器によって取得された771,467データのうち、渋滞が発生している233,246データを使用する。なお、本研究では、対象区間内に30 km/h以下の速度を観測している地点が一箇所でもあれば、渋滞が発生しているデータとする。

この検知器データの速度データから5分毎の $ITT(t)$ 、

$RTT(t)$ 、 $C(t)$ 、 $x(t)$ を算出する。なお、 $RTT(t)$ の算出

にはタイムスライス法を用いる。また、これらの変数の値の算出の際に、速度が0 (km/h)のデータが含まれていることに起因して解析に用いることが適切ではないデータは省く。具体的には、 $ITT(t) = \infty$ であるもの、

$C(t) = 0$ であるもの、 $RTT(t) > 120$ のもの、

$RTT(t) > 5 \cdot ITT(t)$ のデータを除外する。なお、これらのデータは全体の0.01%以下である。

予測モデルのパラメータの推定では、以上で述べたデータのうち、ランダムに抽出された半数データを用いる。残りの半数のデータは、予測値の性能を検証するために用いる。

(2) パラメータ推定方法

$C(t)$ は連続値なので0.1刻みの階級に分け、0.5未満、

0.5以上0.6未満、…、2.0以上の17区分として用いる。最上流部から渋滞末尾までの距離 $x(t)$ は、検知器の速度データをもとに渋滞区間を判別し算出する。なお、ある検知器が30 km/h以下の速度を観測すれば、その検知器の支配区間全体が渋滞しているとみなす。ゆえに $x(t)$ は検知器の数と同じ数だけの場合 (18通り) をもつ離散的な変数になる。

a) 独立を仮定しないモデル

$C(t)$ の階級と $x(t)$ の値の組み合わせ毎にデータを分割し、分割されたデータそれぞれの平均と分散を求めるこ

とで確率密度関数 $p(F_{c,x}|C(t),x(t))$ を求める。

b) 独立を仮定するモデル

$C(t)$ と $x(t)$ が独立と仮定して $F_{c,x}$ の確率分布を求めるために、1変数の確率密度関数である $p(F_c|C(t))$ 、 $p(F_x|x(t))$ を求める。次に、そこで得られた確率密度から $F_{c,x}$ の分布を求める必要があるが、解析的には求めることはできない。そのため、式(2)を用いた MCMC (Markov Chain Monte Carlo methods) による数値シミュレーションを行い、 $F_{c,x}$ の分布を求める。

(3) パラメータ推定結果

a) 独立を仮定しないモデル

図-1に、推定されたパラメータ $F_{c,x}$ の平均値を示す。データの不足により推定できなかった場合は『—』とする。縦軸はIITの変化率 $C(t)$ を表す。したがって、図の上部は5分前に比べてIITが小さくなっていく場合であり、下のほうは5分前に比べてIITが大きくなっていく場合である。横軸は渋滞末尾までの距離 $x(t)$ を表す。図の左のほう为上流から渋滞末尾までの距離が短い場合で、右のほうが長い場合である。ボトルネックの位置が最下流部であるため、渋滞は右から左に延伸する。また、横軸を渋滞長は、右端からのセルの数で見ることができ、左のほうが渋滞長が長い場合で、右のほうが渋滞長が短い場合を示している。

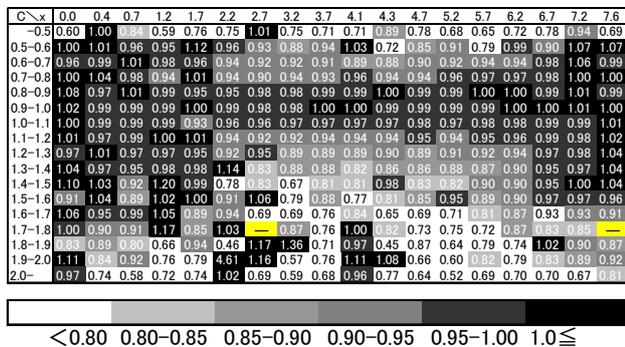


図-1 $F_{c,x}$ の平均値

図より $C(t)$ が1前後の場合、平均値は1に近い値である

ことがわかる。つまり、5分前と比べて交通状態の変化が小さい時は、今後もあまり変化がないと予測される。

$C(t)$ が1.7-1.8では、データが1つもないため予測できない場合が2箇所ある。また、 $C(t)$ が大きな値の時に、隣り合うセル間で推定値が大きく異なる箇所がある。隣り合うセルでは交通状態が似ているため、推定値が大きく異なるとは考えにくく、データ数が少ないために発生したゆらぎだと考えられる。

図-2は、同様に $F_{c,x}$ の10%タイル値と90%タイル値の差を示したものである。この差が大きい場合は、推定された旅行時間の性能が低いことを表し、逆に差が小さいと、旅行時間の性能は高いと見なせる。

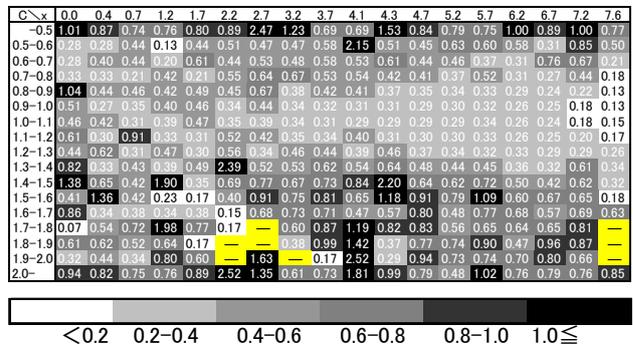


図-2 $F_{c,x}$ の10%タイル値と90%タイル値の差

$C(t)$ が1前後の場合、10%タイル値と90%タイル値の差が小さいため、旅行時間予測値の精度が高いことを示している。 $C(t)$ が1.7-2.0のときに、データがない、もしくは1つしかなく、10%タイル値と90%タイル値の差が計算できない場合があった。この図では示していないが、10%タイル値が負になる場合があり、データ数が不足していることが影響していると考えられる。

b) 独立を仮定するモデル

図-3に、 $C(t)$ と $x(t)$ が独立であると仮定して推定されるパラメータ $F_{c,x}$ の10%タイル値と90%タイル値の平均値を示す。

C\x	0.0	0.4	0.7	1.2	1.7	2.2	2.7	3.2	3.7	4.1	4.3	4.7	5.2	5.7	6.2	6.7	7.2	7.6
-0.5	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.01
0.5-0.6	0.98	0.97	0.94	0.96	0.95	0.96	0.94	0.96	0.97	0.97	0.97	0.97	0.96	0.97	0.97	0.98	0.99	1.00
0.6-0.7	0.97	0.96	0.94	0.96	0.95	0.96	0.95	0.95	0.97	0.97	0.96	0.96	0.96	0.96	0.97	0.98	0.99	1.00
0.7-0.8	0.98	0.97	0.95	0.97	0.96	0.97	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.99	1.00
0.8-0.9	1.00	0.99	0.98	0.99	0.98	0.99	0.98	0.98	0.99	0.99	0.98	0.98	0.98	0.98	0.99	1.00	1.00	1.00
0.9-1.0	1.00	1.00	0.98	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00
1.0-1.1	0.99	0.98	0.97	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.99	1.00
1.1-1.2	0.98	0.97	0.96	0.97	0.96	0.97	0.96	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.98	1.00
1.2-1.3	0.96	0.95	0.94	0.95	0.95	0.95	0.94	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.97	0.97	1.00
1.3-1.4	0.96	0.95	0.92	0.94	0.94	0.94	0.93	0.95	0.95	0.96	0.95	0.95	0.95	0.95	0.96	0.97	0.98	1.00
1.4-1.5	0.97	0.96	0.93	0.96	0.94	0.95	0.94	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.97	0.98	0.99	1.00
1.5-1.6	0.98	0.96	0.93	0.96	0.95	0.95	0.94	0.95	0.97	0.97	0.96	0.96	0.96	0.96	0.97	0.98	0.99	1.00
1.6-1.7	0.95	0.94	0.91	0.93	0.93	0.93	0.92	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.97	0.98	1.00
1.7-1.8	0.95	0.94	0.90	0.93	0.93	0.93	0.92	0.94	0.96	0.95	0.95	0.95	0.95	0.95	0.96	0.97	0.98	1.00
1.8-1.9	0.95	0.94	0.90	0.93	0.93	0.93	0.91	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.98	1.00
1.9-2.0	0.98	0.96	0.92	0.96	0.94	0.95	0.93	0.95	0.97	0.97	0.96	0.96	0.96	0.96	0.97	0.98	0.99	1.00
2.0-	0.90	0.90	0.85	0.88	0.89	0.87	0.88	0.91	0.95	0.95	0.93	0.93	0.93	0.93	0.95	0.96	0.98	1.00

<0.80 0.80-0.85 0.85-0.90 0.90-0.95 0.95-1.00 1.0 ≤

図3 $F_{c,x}$ (独立) の10%タイル値と90%タイル値の平均値

$C(t)$ が大きくて $x(t)$ が小さい場合、つまり、5分前に比べてIITが大きくなる場合で渋滞長が長い場合、その時刻ではIITに比べてRTTが小さくすると推定される。

図4に、 $C(t)$ と $x(t)$ が独立であると仮定して推定されるパラメータ $F_{c,x}$ (独立) の10%タイル値と90%タイル値の差を示す。

C\x	0.0	0.4	0.7	1.2	1.7	2.2	2.7	3.2	3.7	4.1	4.3	4.7	5.2	5.7	6.2	6.7	7.2	7.6
-0.5	0.19	0.18	0.19	0.19	0.20	0.19	0.18	0.18	0.18	0.17	0.17	0.18	0.18	0.17	0.16	0.16	0.16	0.13
0.5-0.6	0.47	0.42	0.45	0.47	0.40	0.48	0.41	0.34	0.33	0.34	0.31	0.31	0.30	0.33	0.27	0.26	0.23	0.15
0.6-0.7	0.40	0.37	0.39	0.40	0.35	0.40	0.36	0.32	0.30	0.32	0.28	0.28	0.29	0.31	0.25	0.25	0.22	0.15
0.7-0.8	0.33	0.33	0.33	0.34	0.31	0.34	0.31	0.28	0.27	0.28	0.26	0.25	0.25	0.28	0.24	0.22	0.21	0.15
0.8-0.9	0.28	0.27	0.28	0.28	0.26	0.29	0.26	0.25	0.24	0.24	0.23	0.23	0.23	0.24	0.22	0.21	0.19	0.14
0.9-1.0	0.24	0.24	0.24	0.24	0.23	0.25	0.23	0.22	0.22	0.22	0.21	0.21	0.21	0.21	0.20	0.19	0.18	0.14
1.0-1.1	0.23	0.24	0.24	0.24	0.23	0.24	0.23	0.22	0.21	0.22	0.21	0.21	0.21	0.21	0.19	0.19	0.17	0.14
1.1-1.2	0.26	0.25	0.26	0.25	0.24	0.26	0.25	0.23	0.22	0.23	0.22	0.22	0.22	0.23	0.21	0.19	0.18	0.14
1.2-1.3	0.30	0.28	0.29	0.30	0.28	0.30	0.28	0.26	0.25	0.26	0.24	0.24	0.24	0.26	0.23	0.21	0.19	0.14
1.3-1.4	0.39	0.36	0.38	0.39	0.34	0.39	0.35	0.31	0.30	0.31	0.28	0.28	0.28	0.30	0.25	0.24	0.22	0.15
1.4-1.5	0.46	0.42	0.45	0.46	0.39	0.46	0.40	0.34	0.32	0.34	0.31	0.30	0.30	0.33	0.27	0.26	0.22	0.15
1.5-1.6	0.47	0.43	0.45	0.46	0.39	0.46	0.41	0.35	0.33	0.34	0.31	0.30	0.30	0.34	0.27	0.26	0.23	0.15
1.6-1.7	0.46	0.41	0.43	0.45	0.38	0.46	0.38	0.34	0.32	0.34	0.31	0.30	0.30	0.33	0.27	0.25	0.22	0.15
1.7-1.8	0.49	0.43	0.46	0.48	0.41	0.49	0.42	0.35	0.33	0.34	0.31	0.31	0.31	0.35	0.28	0.27	0.23	0.16
1.8-1.9	0.49	0.45	0.47	0.48	0.41	0.49	0.42	0.35	0.33	0.35	0.31	0.31	0.31	0.34	0.28	0.26	0.23	0.16
1.9-2.0	0.55	0.48	0.51	0.53	0.44	0.54	0.45	0.37	0.35	0.37	0.33	0.32	0.32	0.36	0.29	0.27	0.23	0.16
2.0-	0.51	0.46	0.50	0.51	0.42	0.52	0.44	0.36	0.35	0.36	0.32	0.31	0.31	0.35	0.28	0.26	0.23	0.16

<0.2 0.2-0.4 0.4-0.6 0.6-0.8 0.8-1.0 1.0 ≤

図4 $F_{c,x}$ (独立) の10%タイル値と90%タイル値の差

$x(t)$ が小さくなるにつれて10%タイル値と90%タイル値の差が大きくなる傾向がある。つまり、渋滞長が長いときのほうが、旅行時間予測値の性能が下がるということである。この原因として、自由流に比べて渋滞流のほうが、速度の変化が大きいことが考えられる。

(4) 独立を仮定するモデルとしないモデルの比較

a) 平均値の比較

第3節a項, b項で示した独立性を仮定しないモデルと仮定するモデルの結果ともに、 $C(t)$ の値が1よりも大きい場合、 $F_{c,x}$ の平均値が小さく推定される。5分前に比

べてIITが小さくなると、その時刻のIITに比べてRTTが大きくなるのが予測され、5分前に比べてIITが大きくなると、その時刻のIITに比べてRTTが小さくなるのが予測される。これらは渋滞の疎密により、速度が速くなったり遅くなったりすることが影響していると考えられる。

独立を仮定しないモデルに比べて、独立を仮定するモデルは $F_{c,x}$ の平均値が全般的に1に近い。これは、独立を仮定した式(2)がベイズの定理を満たしておらず、もとの分布 $p(F)$ の分布の形を引きずっていることが原因と考える。なお、 F は、パラメータ推定に用いた過去のデータ全体の平均値と分散をパラメータとしても正規分布に従う確率変数である。独立を仮定する場合、ベイズの定理では、

$$p(F_{c,x}|C(t),x(t)) = \frac{p(F_c|C(t)) \cdot p(F_x|x(t))}{p(F)} \quad (3)$$

が成り立つ。この式で示されるように、理論的には $p(F)$ で除する必要がある。しかし、 $p(F)$ で除すとこの分布に正規分布を仮定しているため、旅行時間の予測値としては不適切な予測値を算出しやすいモデルになってしまうことがある。具体的には、 $p(F_c|C(t))$ と $p(F)$ または $p(F_x|x(t))$ と $p(F)$ の最頻値の位置が離れると、 $p(F)$ が極めて小さな値をとるときに $p(F_{c,x}|C(t),x(t))$ が大きくなる。つまり、パラメータ F が負の値や5以上のときに式(3)の値が大きくなってしまふ。

b) 分散の比較

独立を仮定するモデルでは、 $x(t)$ が小さくなるにつれて分散が大きくなる傾向がある。 $C(t)$ の値が1前後の場合には独立を仮定しないモデルでも同じ傾向を示している。しかし、 $C(t)$ の値が極端に大きい場合や小さい場合は、データ数が少ないため、不安定な挙動を示していた。データ数がたくさんあり、独立性を仮定しないモデルでも予測できた箇所は、独立性を仮定するモデルでも同じ傾向を示すことができる。

独立性を仮定しないモデルでは、分散が大きくて10%タイル値が負になる場合があるが、独立を仮定するモデ

ルでは、10%タイル値と90%タイル値の差が最大でも0.55であり、10%タイル値が負になるような場合はない。本研究で用いたデータセットの場合では、独立を仮定しないモデルでは、予測値の性能を推定できない場合があるのに対して、独立を仮定するモデルでは、どの場合でも推定可能であった。

c) モデル性能の比較

10%タイル値と90%タイル値の間に入る確率の理論値は80%なので、予測は80%の確率で当たることが想定される。第2章2節で述べたように、同じ確率で当たる予測モデルが複数あり、予測値の幅が異なっているとき、幅の狭いモデルのほうが良いモデルだといえる。つまり、当たる確率が一定のとき、予測幅の大きさとモデルの性能を調べることができる。

実際に予測が当たる確率は独立を仮定しないモデルでは87%であった。理論よりも幅を広く推定している原因として、データが少ない場合が考えられる。データ数が少ないと分散が大きくなり、区間推定の幅が広がるからである。一方、独立を仮定するモデルでは76%であり、独立を仮定しないモデルよりも理論値に近い結果となった。

4. 結論

本研究では、Weiらの旅行時間予測モデルを、点推定するモデルから区間推定するモデルへ拡張した。さらに、Weiらのモデルで用いられていた条件変数 $C(t)$ と $x(t)$ が独立であると仮定してモデルを拡張した。

区間推定への拡張により、旅行時間の推定値と同時に、その予測結果の性能を推定できるようになった。第4章の分析では、性能を推定することにより、データ数が少ない場合は予測値の性能が下がることの確認などに役立った。

独立性を仮定したモデル化により、過去のデータが少ない場合でも予測できることが期待できる。この手法により、条件変数として新たなパラメータを導入することができる。今後、条件変数の数を増やして、なるべく狭い幅で区間推定できるようになれば、旅行時間予測値の性能を向上できると期待される。

4章の分析からは、サンプル数が多く、独立を仮定しないモデルでも推定できる部分の傾向を示すことができた。具体的には、5分前に比べてITTが大きくなると、同時刻のRTTがITTに比べて小さく予測されることや、渋滞長が長くなると予測値の性能が落ちることを示せた。しかし、独立を仮定するモデルは、 F の分布系が正規分布であると仮定しているため、ベイズの定理を満足した解析が行うことができない。その影響により、独立を仮定する方法を用いても追加できる条件変数の個数には

限りがあることがわかった。

付録

$C(t)$ と $x(t)$ が独立と仮定するモデルのパラメータ推定を行うために、1変数を条件とする確率密度関数が必要とされた。 $C(t)$ と $x(t)$ をそれぞれ変数とする F_c 、 F_x の平均と分散を図-5と図-6に示す。

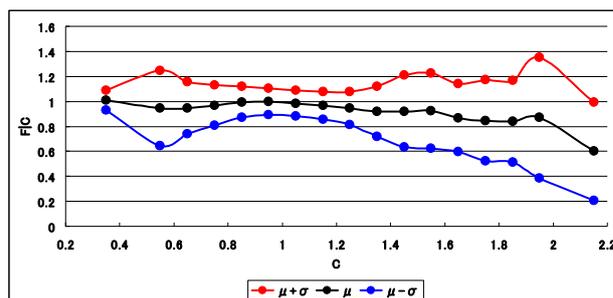


図-5 F_c の平均と分散

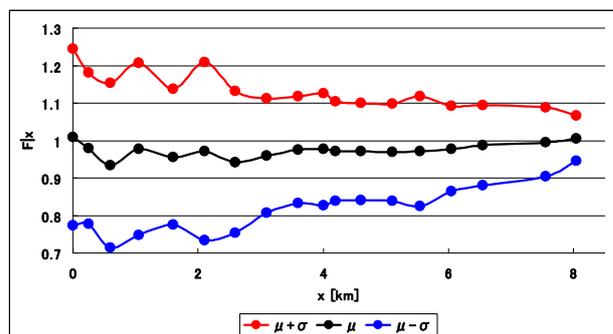


図-6 F_x の平均と分散

$C(t)$ の値が1を超えている場合、つまり、5分前に比べてITTが大きくなっている場合、その時刻のITTに比べてRTTは小さく予測されることを示す。 $x(t)$ が小さい場合、すなわち渋滞長が長い場合、分散が大きく予測されることを示している。

参考文献

- 1) 舟橋賢二, 西村茂樹, 堀口良太, 赤羽弘和, 桑原雅夫, 小根山裕之: VICS 蓄積データを用いた旅行時間

短期予測手法に関する研究, 第 27 回土木計画学研究
発表会講演集, No.27, CD-ROM, 2003.

- 2) 宗像恵子, 田村勇二, 割田博, 白石智良: 首都高速
道路におけるリアルタイム予測シミュレーションの
開発, 第 29 回交通工学研究発表会論文集, pp.293-
296, 2009.

Wei, C., Iryo, T. and Asakura, Y. : Implementation of Short-
Term Travel Time Prediction Model on Urban Expressway,
International Journal of ITS Research, 7(2), pp. 101-108., 2009.

(2011.8.5 受付)

STATISTICAL ANALYSIS ON PERFORMANCE OF TRAVEL TIME PREDICTION

Katsuya SAKAI , Chong WEI , Takahiko KUSAKABE and Yasuo ASAKURA