

特徴空間射影を用いた交通情報予測手法

熊谷 正俊[†] 伏木 匠[†] 佐野 豊[‡] 鈴木 研二[‡] 横田 孝義[‡]

[†]株式会社日立製作所日立研究所 〒319-1292 茨城県日立市大みか町 7-1-1

[‡]株式会社日立製作所情報制御システム事業部 〒319-1293 日立市大みか町 5-2-1

E-mail: [†]{mkumagai, tfushiki, tyokota}@gm.hrl.hitachi.co.jp, [‡]{yutaka_sano, kenji-g_suzuki}@pis.hitachi.co.jp

あらまし 交通情報時系列データを主成分分析を用いて特徴空間に射影し、特徴空間における射影点座標を曜日、季節などの日種の組み合わせを説明変量として近似する推定関数を得る。交通情報の予測演算では、予測対象日の日種情報から前記推定関数を用いて特徴空間における予測座標を計算し、これを時系列データ空間に逆射影することで、時系列データの予測値を得る。本手法の特徴は、任意の日種の組み合わせのもとで、任意の時間分解能を持つ交通情報の予測を可能とすることである。

キーワード 旅行時間予測, 特徴空間, 主成分分析

Traffic Information Prediction Method Based on Feature Space Projection

Masatoshi KUMAGAI[†] Takumi FUSHIKI[†] Yutaka SANNO[‡] Kenji SUZUKI[‡]
and Takayoshi YOKOTA[†]

[†]Hitachi Research Laboratory, Hitachi, Ltd. 7-1-1 Omika, Hitachi, 319-1292 Japan

[‡]Information & Control Systems Div., Hitachi, Ltd. 5-2-1 Omika, Hitachi, 319-1293 Japan

E-mail: [†]{mkumagai, tfushiki, tyokota}@gm.hrl.hitachi.co.jp, [‡]{yutaka_sano, kenji-g_suzuki}@pis.hitachi.co.jp

Abstract We propose a prediction method for traffic information with any period (e.g., travel time of every five minutes), which reflects a combination of some day types (day, season, vacation, weather, and so on). Firstly, we project the past traffic information data to a feature space, using the principal component analysis method. Secondly, we get an estimate model which approximate feature values as a function of day types. When predicting traffic information, the estimate model gives an estimated feature value using day types in a future day, and predictive information is given by inverse projection of the feature value.

Keyword Travel Time Prediction, Feature Space Projection, Principal Component Analysis Method

1. 緒言

2002年6月に施行された道路交通法の改正により、道路交通情報を収集、加工して提供するサービス事業への門戸が民間企業に対して開かれた。また、各都道府県警察が整備を進めている交通管制システムから得られるデータについて、統計的利用の環境が整いつつある。

本論文で報告するのは、こうした交通管制システムから、あるいは独自に収集して得られる旅行時間情報を統計的に加工して、中長期の旅行時間情報の予測に活用する手法である。旅行時間の予測情報を提供する上では、

- ・ 1日の旅行時間の変化を利用者が認識可能であるように数分～1時間程度の分解能を持つ情報を、
- ・ 曜日や季節など予測対象日の条件をきめ細かく反映して予測する、

という点を重視しており、本研究では旅行時間情報を特徴空間に射影し、特徴空間上で予測を行うことで、それらを実現している。本論文の2章では旅行時間の中長期予測を行

う上での課題について述べ、3章にて提案する旅行時間予測手法を解説し、4章はその検証結果である。

2. 旅行時間中長期予測の課題

旅行時間の中長期予測(1日～1年先程度の旅行時間を予測)は、同一の事象の下では類似の交通状況が再現されるという観測経験に基づく考え方に立脚している。そして、中長期予測において提供する情報の分解能としては、数分から1時間程度の時間を単位として、いずれの時間帯にどれだけの旅行時間を要する傾向にあるか、それが判別可能であることが求められる。

上述の考え方を如実に反映した予測手法としては、数量化I類が一般的である^{[1][2]}。この方法は、予測対象量に影響を及ぼしていると考えられる曜日、十五日、季節、学校休業期間、天気などの因子を表す説明変量の線形結合によって予測値を計算するものである。これは、予測対象量を各因子に由来する変動量の合成によって表す方法と言える。

ただし、説明変量を上記日単位の情報とした場合には、予測対象量も日単位の情報に限られる。1日あたりの地点交通量、ある特定の時間帯の旅行時間などである。

一方、分解能に対する要求に応える方法としては、交通情報を事象に応じて1日ごとに分類し、分類単位ごとに同時刻平均値などの代表値をもって予測値とする方法が一般的である^[3]。これは、曜日などの因子によって1日ごとの交通情報時系列データをラベリングしたデータベースを構成し、データベース検索により必要とする情報を取り出す方法であると言える。この方法によれば、周期的に変化する交通状況の最小周期である日単位の情報をラベルとして、任意の時間間隔の交通情報を取り扱うことができる。しかし、この方法は因子の組み合わせによってデータの分類を行うため、曜日、季節、五十日、天気、学校休業期間など複数の因子が関係した情報として交通情報を取り扱う際には、データベースの分類単位が各アイテム(曜日、天気などの項目)を構成するカテゴリ(曜日であれば月曜日、火曜日などの項目値)の積に比例して増大する。それゆえ、このような分類方式には、アイテム数が多岐に亘る場合、

- ・ 分類単位ごとのサンプル数が少なく、統計的有意性が低下する
- ・ 予測対象日の事象がデータベースから検索できるとは限らない

という問題がある。後者は具体的には、予測対象日が例えば春の学校休業期間中の金曜日の五十日であった場合に、交通情報をデータベースに蓄積した期間にそれと同一事象下でのデータが存在しない確率が、データベースの細分化に伴って増大するという問題である。

先に述べたように、交通現象は1日を最小単位として周期的に繰り返される現象であり、その予測においては、曜日、平日/休日、祝祭日、五十日、季節、天気などの因子が重要な役割を果たすと考えられる。よってこれらの因子を、1日単位の交通情報に影響を与える情報として日種情報と呼ぶことにする。日種を用いた交通情報予測の一般的な手法には上述の問題があり、次章においては、かかる問題を解決する予測手法について述べる。

3. 特徴空間射影による旅行時間予測手法

3.1. アルゴリズムの概念

前章で述べた日種を説明変量とする数量化I類による手法は、日単位の情報を予測対象量としている限りにおいては、きわめて合理的な方法である。ならば、1日分の時系列データを代表する日単位の情報を予測対象とすることができれば、数量化I類など日単位情報の予測に適した手法を、任意間隔の時系列データの予測に適用可能であると考えられる。

一方、データベースに格納された時系列データを用いて予測を行う手法に前述の問題があるのは、予測情報の生

成に分類と検索という手段を用いているためと考えられる。これに代えて、既存の時系列データの合成により予測情報を生成する手段をとれば、この問題は回避される。そのような合成の対象となりうる時系列データのひとつに、直交基底関数がある。複数の直交基底関数は互いに独立であり、その線形和により任意の時系列データを近似的に表現することができる。例えば、フーリエ変換における直交基底関数は周期の異なる複数の三角関数であり、フーリエ変換はそれら三角関数の合成により、任意の周期信号を近似する。ただし、フーリエ変換は周期信号を表現するのに適切な方法ではあるが、周期的とはいえ、周期ごとに異なる波形を持つ交通情報のような時系列データを表現するには必ずしも適当ではない。

そこで、本研究においては、旅行時間時系列データを適切に表現可能な基底データを得る手段として、1日ごとの時系列データをサンプルとする主成分分析を用いた。主成分分析によって時系列データから得られる結合係数は、時系列データ空間における各主軸のベクトルであり、主成分得点を係数とする線形結合により、任意サンプル日の時系列データを表現することができる。各主軸は、時系列データ空間におけるサンプルの分散が最も大きい方向から順に得られる独立なベクトルであり、主軸の全てではなく、分散の大きい順にいくつかの主軸を選んで時系列データを再構築した場合、2乗誤差最小で元データを近似するベクトルのセットである。よって、結合係数は旅行時間時系列データを表現する基底データとして適当な性質を持つと言える。また、主軸上での各サンプルの座標を表す主成分得点は、主軸ごとに独立であり、かつ、日単位の情報である。つまり、各主軸上での主成分得点のそれぞれは、先に述べた1日分の時系列データの代表値として、日種を説明変量とした予測の対象として扱うことが可能である。そして、予測された主成分得点を係数として基底データである結合係数を合成することで、1日分の時系列データの予測値を得ることができる。

この考え方は、次のような見方でもとらえることもできる。主軸で構成される空間を、旅行時間時系列データを射影した特徴空間とみなしたとき、主成分得点は各サンプル日の時系列データの射影点を表す特徴量である。また、特徴量の予測値を用いて結合係数を合成する処理は、特徴空間における予測値から時系列データ空間への逆射影である。すなわち、上記手法は、時系列データ空間内で適当な処理の難しい情報を特徴空間に射影して扱い、それをまた時系列空間に逆射影することにより、目的とする予測データを得る手法であると言える。なお、時系列データを特徴空間に射影する際に主成分分析を用いることは、先に述べたように情報の縮退による誤差が最小であるという意味で最尤である。また、特徴空間において、特徴量の実績値から日種を説明変量とする予測モデルを導出することと、かかる予測モデルを用いた特徴量予測値の演算に数量化I類を用いることは、

線形予測モデルとしては最尤推定である。よって、主成分分析と数量化1類により特徴空間で時系列データの予測をすることは、線形の枠組みでは最適な手法と考えられる。

3.2. アルゴリズムの詳細

前節に述べた交通情報の予測手法について、以下、詳細を述べる。

1日分の旅行時間時系列データを、行ベクトル

$$\mathbf{x}_i^* = [x_{i1}^* \ x_{i2}^* \ \dots \ x_{iM}^*] \quad (1)$$

とする。添え字 i は日付を表し、 M は1日あたりのデータの数である。時系列上の各データは等間隔にサンプリングされたものであり、たとえばサンプリング周期を5分とすると、 $M=288$ である。交通情報の蓄積期間 N 日間の時系列データから、 \mathbf{x}_i^* の各要素について同時刻平均 $\bar{\mathbf{x}}$ を

$$\bar{\mathbf{x}} = \left[\frac{1}{N} \sum_i x_{i1} \quad \frac{1}{N} \sum_i x_{i2} \quad \dots \quad \frac{1}{N} \sum_i x_{iM} \right] \quad (2)$$

により計算し、 $\bar{\mathbf{x}}$ に対する \mathbf{x}_i^* の偏差

$$\mathbf{x}_i = [x_{i1} - \bar{x}_1 \quad x_{i2} - \bar{x}_2 \quad \dots \quad x_{iM} - \bar{x}_M] \quad (3)$$

を得る。これを N 日間について並べたものを、

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} \quad (4)$$

とする。 \mathbf{X} の列毎の共分散行列を \mathbf{V} とすると、これを固有値分解して得られた各固有ベクトル

$$\mathbf{w}_j = [w_{1j} \quad w_{2j} \quad \dots \quad w_{Mj}]^T \quad (5)$$

が結合係数であり、対応する固有値の降順にそれらを並べた行列

$$\mathbf{W} = [w_1 \quad w_2 \quad \dots \quad w_M] \quad (6)$$

が結合係数行列である。 \mathbf{W} から、主要な結合係数を取り出した

$$w_1, w_2, \dots, w_p \quad (p < M)$$

がここで用いる基底データであり、これを基底行列

$$\mathbf{P} = [w_1 \quad w_2 \quad \dots \quad w_p] \quad (7)$$

と表記する。 \mathbf{P} に対して、

$$\mathbf{T} = \mathbf{XP} \quad (8)$$

で得られる \mathbf{T} が主成分得点行列であり、その要素 t_{ij} が、 i 日の時系列データを特徴空間に射影したときの第 j 主軸における特徴量に相当する。

ここまでが旅行時間の実績値から特徴量を求める処理であり、続いて、特徴量を日種の関数で表現する方法について述べる。主成分得点行列 \mathbf{T} から第 j 列を取り出し \mathbf{t}_j とする。

また、交通情報を蓄積した N 日間における日種の情報を行列 \mathbf{D} で表すことにする。日種行列 \mathbf{D} は、その要素 d_{ij} が i 日における日種 j への該当を、1(該当する)と0(該当しない)で表す。本研究では、特徴量を日種情報で予測するのに数量化1類の手法を用いる、つまり予測関数の関数形を日種の線形結合としてしているので、その係数を予測係数ベクトル \mathbf{a}_j とすると、

$$\mathbf{t}_j \cong \mathbf{D}\mathbf{a}_j \quad (9)$$

という関係が成立し、上式の2乗誤差を最小化するという意味での最尤予測係数は、擬似逆行列 \mathbf{D}^+ を用いて、

$$\mathbf{a}_j = \mathbf{D}^+ \mathbf{t}_j \quad (10)$$

と表される。これを各主軸上の特徴量について求めたものが、予測係数行列

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_p] \quad (11)$$

である。なお、 \mathbf{t}_j の平均値が0であることから、定数項は存在しない。

以上で求めた基底行列 \mathbf{P} と予測係数行列 \mathbf{A} を用いて予測対象日の旅行時間時系列データを予測する処理は下記のようになる。まず、予測対象日の日種情報を、日種行列 \mathbf{D} の各行と同様の表記法で記述した行ベクトル \mathbf{d} で表す。日種情報 \mathbf{d} に対し、特徴量の予測値

$$\mathbf{u} = [u_1 \quad u_2 \quad \dots \quad u_p] \quad (12)$$

は、

$$\mathbf{u} = \mathbf{d}\mathbf{A} \quad (13)$$

である。かかる特徴量予測値は基底データを合成する際の係数としての性質を持つ。ゆえに、旅行時間の予測値 \mathbf{y} は、基底行列 \mathbf{P} を用いて、

$$\mathbf{y} = \mathbf{u}\mathbf{P}^T + \bar{\mathbf{x}} \quad (14)$$

により得られる。

以上を模式的に表すと図1のようになる。まず、旅行時間時系列データ \mathbf{x}_i^* を解析して基底行列 \mathbf{P} と、対応する特徴量 \mathbf{T} を求める。基底行列 \mathbf{P} と特徴量 \mathbf{T} は、その積と $\bar{\mathbf{x}}$ との和によって任意の \mathbf{x}_i^* を近似表現可能な行列である。次に、特徴量 \mathbf{T} と日種行列 \mathbf{D} から式(10)(11)により予測係数行列 \mathbf{A} を得る。予測係数行列 \mathbf{A} と予測対象日の日種ベクトル \mathbf{d} から特徴量の予測値 \mathbf{u} を求め、これを用いて基底行列 \mathbf{P} を線形合成することで、旅行時間の予測値 \mathbf{y} を得る。

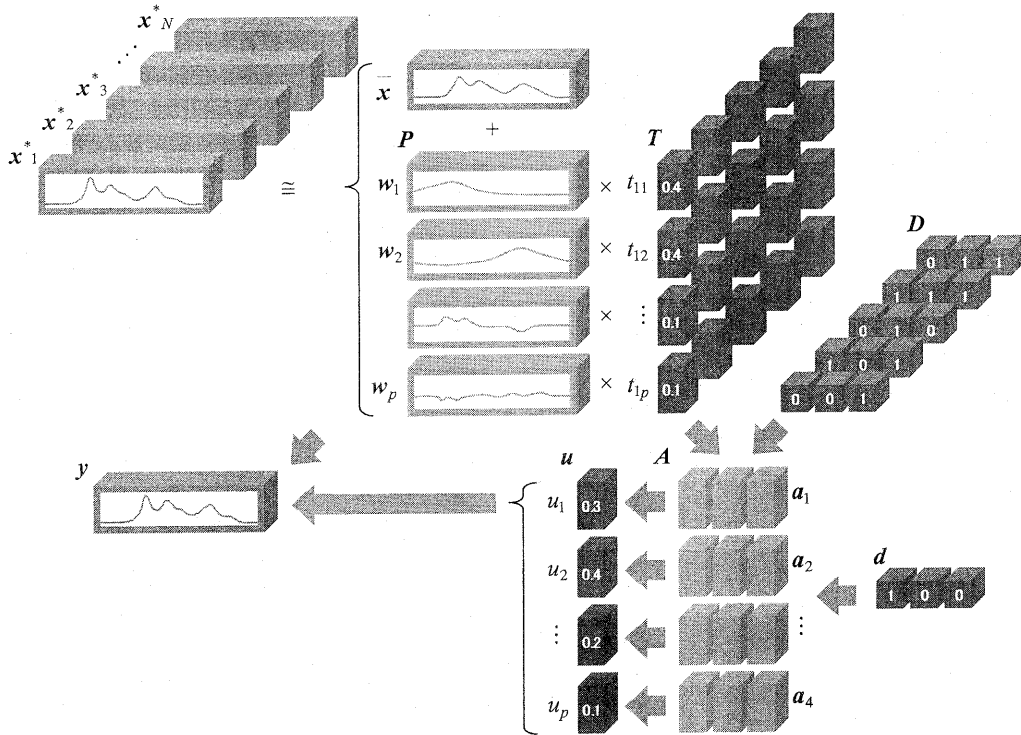


図 1 予測手法の模式図

4. 予測手法の評価

4.1. 評価の方法

以上に述べた旅行時間予測手法を、交通管制システムの旅行時間データを用いて評価した。評価に使用したデータは2001年度のものであり、5分間隔にリンク旅行時間が記録されたものである。国道16号線、横浜町田IC近傍の上り線1172[m]の区間を評価区間とした。上記1年間のデータから、標本データと評価データを重複させることなく1年間にわたる均一な評価を行うために、以下の処理により評価を実施した。

- (1) 2001年度の旅行時間データから、偶数月(4月, 6月, ..., 2月)のデータを取り出し、異常値除去などの前処理を施した上で標本データとする。
- (2) 標本データをもとに、前章に述べた基底行列 P と予測係数行列 A を算出する。
- (3) 奇数月(5月, 7月, ..., 3月)の日種情報をもとに、(2)で求めた基底行列 P と予測係数行列 A を用いて、奇数月における旅行時間予測データを算出する。
- (4) (3)で得た旅行時間予測データを、奇数月の旅行時間実績データと比較評価する。

以上により、本報告では標本データと評価データを分離している。標本データから時間的に遡って予測データを作成する期間もあるが、アルゴリズムの妥当性を評価する上での問題はない。もちろん、実際に本手法を予測に用いる場合には、2003年度のデータを標本データとして2004年度のデータを予測するといった形になる。なお、評価用のプログラムは豊富な行列計算機能を持つMaTX^[4]を用いて作成した。

図2～図5に評価の結果をグラフで示す。図2は、2001年度5月の週末を挟んだ4日間における、本手法による旅行時間予測データと、実績データのグラフである。いずれも、5分間隔の時系列データを24時間分プロットしている。ここでは、予測に用いる日種情報として、曜日、祝祭日、五十日、季節などの情報を用いた。予測データを計算するにあたり、標本データには異常値除去などの前処理を施したが、図2に示したグラフでは、計算された予測データ、および実績データに対して平滑化などの処理は施していない。これらのグラフから予測データは次の性質を持つことがわかる。

- (1) 短周期変動のある実績データに対して、予測データはそれらをフィルタリングしたようなデータとなる。

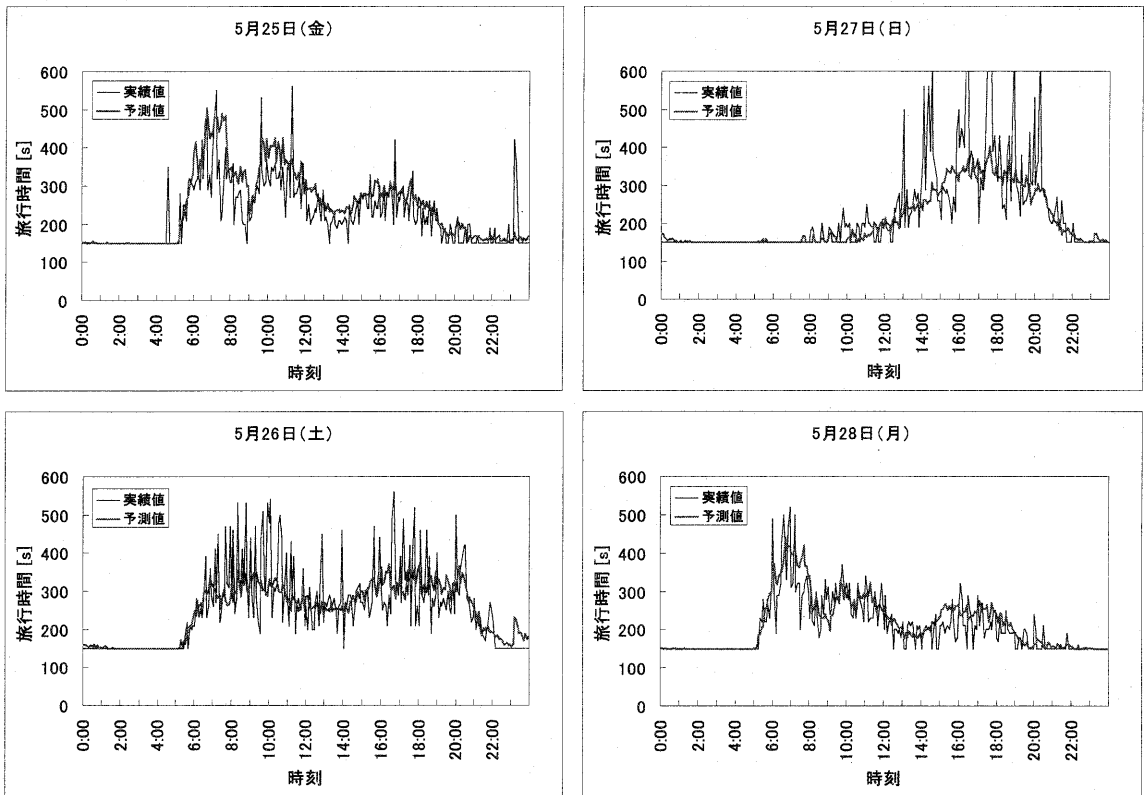


図 2 旅行時間予測値と実績値の比較

- (2) 実績データに見られる朝, 昼, 夕方ピークを反映した予測データが得られる。
 - (3) 曜日の変化, 特に平日と土曜日, 日曜日の違いも顕著に反映されている。
- (1)のフィルタリングの効果は, 寄与率の低い主軸として, ノイズ成分が除かれたものと考えられる。

図 3のグラフは, 10時台における実績データに対する予測データの誤差率の分布を, 累積度数で示したものである。横軸には実績データに対する予測データの誤差率階級がとられている。縦軸は, その累積度数を正規化したものである。ここでは1日あたり288点のデータがあり, それが奇数月全てで184日分あるため, $288 \times 184 = 41472$ 点について誤差率の分布を求めた。なお, この評価では実績データと予測データのそれぞれにおいて急峻な変化の影響を取り除くため, 前後15分の区間(7点)で移動平均処理をした。移動平均処理による平滑化の効果は, たとえば図 2における5月28日のデータについては図 4のようなものである。

図 3の度数分布表からは, たとえば誤差率20%の範囲

には, 全体の約7割が含まれており, 誤差率30%の範囲には, 8割強が含まれていることがわかる。これは, 10時台に当該区間を走行する場合に, 10回に8回は, 実所要時間の $\pm 30\%$ の範囲の予測値が得られるであろうことを意味する。このような誤差分布を, 6時台から23時台まで評価した結果が図 5のグラフである。これを見ると, 朝の8時台, 夜の19時台前後で誤差が増大していることがわかるが, 最も誤差の大きい8時台において誤差率30%範囲の累積度数が約0.65, 19時台と20時台に同じく0.75弱である他は, 誤差率30%範囲の累積度数が0.8以上である。利用者の60%がリアルタイム所要時間情報において30分の提供時間に ± 5 分($\pm 17\%$)の精度を求めているという既存のアンケート結果^[5]に基づき, 首都高速道路公団の割田らが長期予測情報では $\pm 33\%$ の精度を目安として採用している観点からすれば^[3], この結果は, 渋滞が特に著しい時間帯を除いては, 10回の走行のうち8回は適切な予測情報を提供可能であることを意味している。

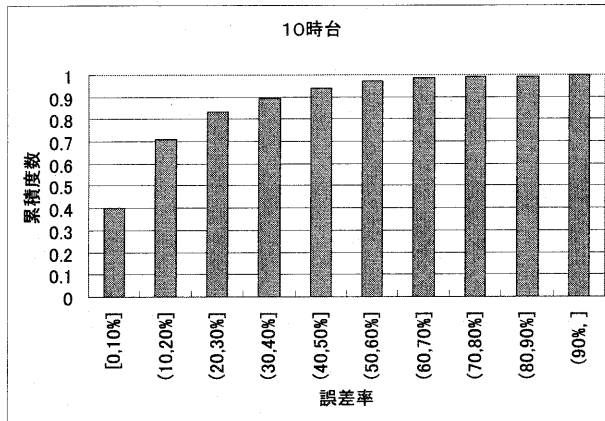


図 3 誤差率の累積度数分布(10 時台)

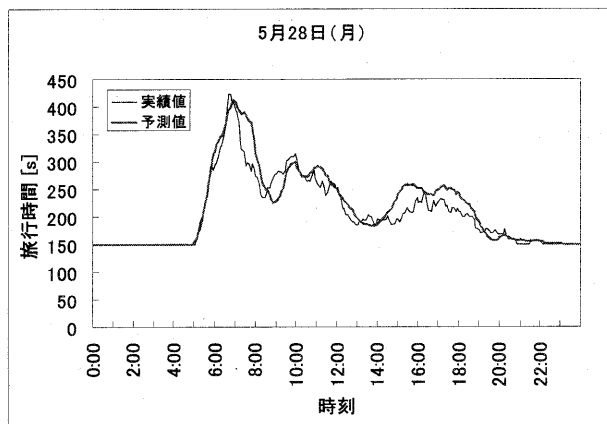


図 4 旅行時間予測値と実績値の平滑化

5. 結言

本論文では、特徴空間射影を用いることで、曜日、季節、五十日などの日種を複合的に関連付けて旅行時間の予測を行う手法について述べ、その有効性を検証した。今後はより広域的な精度評価と、それに基づくアルゴリズムの改良を進めるとともに、リアルタイムに受信される旅行時間情報で予測値を補正する手法の開発を進める考えである。また、本報告では特徴空間での予測に数量化I類を用いたが、これをニューラルネットなどの非線形モデルで置き換えることで予測精度の向上が図れるか、その効果の検討も今後の課題のひとつである。

文 献

- [1] 澤, 交通計量経済学(改訂版), pp.37-40, 成山堂書店, 2000.
- [2] 横井, 児玉, 星, “交通渋滞予測手法に関する一考察,” 第16回交通工学研究発表会論文報告集, pp.69-7, Nov. 1995.
- [3] 割田, 吉田, “首都高速道路における所要時間変動特性の分析,” 第22回交通工学研究発表会論文報告集, pp.61-64, 2002-11.
- [4] <http://www.matx.org/>
- [5] 首都高速道路公団, “将来管制システムに関する研究(平成11年度),” Feb. 2000.

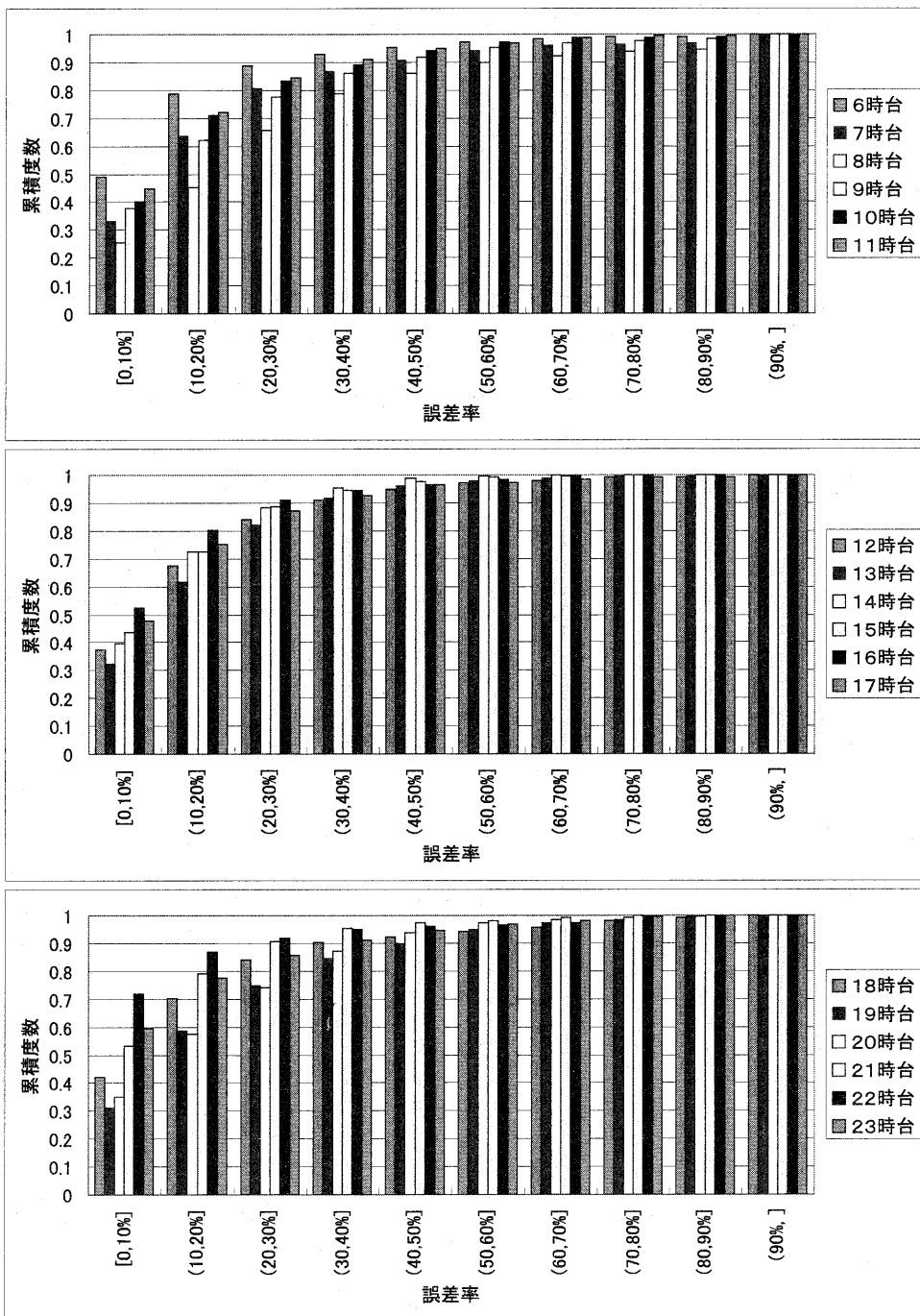


図 5 誤差率の累積度数分布