

特徴空間軌跡を用いた交通渋滞の動的予測

熊谷 正俊 蛭田 智昭 谷越 浩一郎 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

E-mail: {masatoshi.kumagai.ws, tomoaki.hiruta.dp, koichiro.tanikoshi.uw, takayoshi.yokota.py}@hitachi.com

あらまし 疎らなプローブカーデータを用いて交通渋滞を動的に予測する技術について述べる。プローブカーは交通情報の収集において効果的な手段だが、その台数の制約により、収集されるデータは多大な欠損を伴う。一方、従来の渋滞予測手法は、欠損のない路上センサデータの使用を前提としており、欠損を伴うプローブカーデータの利用には適さない。この課題の解決のため、本研究では特徴空間射影を用いる渋滞予測手法を開発した。これは、プローブカーデータの空間的欠損を補う技術として、以前に報告した推定補完手法の拡張であり、特徴空間上の射影点軌跡をトレースすることで予測情報を算出する。本報告は開発の第一報であり、予測アルゴリズムの妥当性を検証した結果について述べる。

キーワード 交通情報, 予測, プローブカーデータ, 欠損値補完, 主成分分析

Dynamic prediction of traffic congestion using feature space trajectory

Masatoshi KUMAGAI Tomoaki HIRUTA Koichiro TANIKOSHI and Takayoshi YOKOTA

HITACHI Ltd. Hitachi Research Laboratory 7-1-1 Omika, Hitachi-shi, Ibaraki, 319-1292 Japan

E-mail: {masatoshi.kumagai.ws, tomoaki.hiruta.dp, koichiro.tanikoshi.uw, takayoshi.yokota.py}@hitachi.com

Abstract This paper discusses a dynamic prediction method of traffic congestion using sparse floating car data (FCD). Floating cars are effective way to collect traffic information; however, because of the limitation of the number of floating cars, there is a large amount of missing data with FCD. On the other hand, traditional prediction methods are designed on the purpose of utilizing road-side sensors with no missing data. Therefore, traditional methods cannot work with sparse FCD. In an effort to address this problem, we developed a new prediction method based on feature space projection. The method is the extension of a spatial imputation method which we previously proposed for floating car systems, and prediction is achieved by tracing projection trajectory in the feature space. This paper is an interim report of the development, and we started testing the algorithm by evaluating prediction accuracy independently of imputation process.

Keyword Traffic information, Prediction, Floating car data, Missing data imputation, Principal component analysis

1. 緒言

道路上の交通状況を計測するには、従来は超音波感知器やループコイル、画像センサを用いるのが一般的であった。いずれも路上に設置された固定センサであり、インフラコストが掛かるため、そのカバーエリアは主要道路に限られている。これに対して、GPSの普及と無線通信の発達が可能としたインフラレスの情報収集手段が、プローブカーシステムである。

プローブカーは道路上を走行する自動車そのものがセンサであり、位置、速度などの車両データを、無線通信を介して交通情報センタにアップリンクする。交通情報センタでは、多数の車両から収集されたデータを、道路上の速度や旅行時間を表すデータに加工し、交通情報として配信する。プローブカーとして専用の車両を走らせる場合もあるが、自家用車、タクシー、ロードサービス車両などを計測車両として活用するのが一般的である^{[1][2]}。

プローブカーの利点は、新たに路上インフラの整備

を必要とせず、GPSを備えた車両、無線通信回線、情報を集約する交通情報センタがあれば、交通情報の収集が可能となる点にある。一方、実用の交通情報サービスに供する上では、プローブカーに特有の問題もある。それは、情報収集が車両の分布に依存する点である。

たとえば、日本全国に10万台のプローブカーが走行していたとしても、交通情報の更新頻度は、平均で1時間に1回程度にすぎない^[3]。これは、一般的な路上センサの情報更新頻度が5分に1回であることと比較すると、非常に疎らなデータであり、交通状況の変化に追従できるデータ密度とは言い難い。それゆえ、プローブカーを交通情報サービスに用いたとしても、交通情報がいつまで経っても更新されない、いつも固定的な経路案内がなされるといった状況は、容易に生じうる。

この問題の解決には、プローブカー台数そのものを増やすという直接的な手段もあるが、情報処理の工夫による解決も可能である。その一つが、筆者らが先に

提案した特徴空間補完である^[4]。

特徴空間補完は、過去のプローブカーデータ履歴の主成分分析によって特徴空間を構成し、特徴空間に欠損を伴う現況データを射影/逆射影することで、欠損データをリアルタイムに推定補完する技術である。道路ネットワークの交通状況を多次元の特徴空間で表現し、現況の観測データから特徴空間座標を動的に決定することで、特徴空間の持つ関連情報に基づいた高精度な推定を可能とする。

一方、交通情報サービスにおいては、現況のみならず、未来の交通状況に関する予測も重要である。しかしながら、前述の特徴空間補完は現況データの補完を行うための技術であり、未来の交通状況を予測するものではない。また、既存の予測手法は欠損の無い時系列データの入力を前提としているものが多く、路上センサデータに基づく予測では効果を発揮するが、欠損を伴うプローブカーデータには対応していない。

この課題に対して、欠損を伴う現況データを反映して、未来の交通状況を動的に予測するには、大きく分けて下記3つのアプローチが考えられる。

- (1) 特徴空間補完の結果を既存の予測手法の入力とする。既存技術の組み合わせであり、シンプルなアプローチではあるが、補完演算による誤差に予測演算の誤差が重畳されることになり、高い精度を期待することは難しい。
- (2) 従来手法の1つであるパターンマッチング予測^[1]を、リンク単位(主要交差点間の道路区間)から、エリア単位(複数のリンク群)に拡大して用いる。すなわち、現況のプローブカーデータが収集されたリンクの交通情報をキーとして、過去の交通情報履歴を検索する。しかし、検索対象である過去の交通情報も欠損を伴うため、プローブカーデータの欠損率が80%程度になると検索のヒット率が著しく低下し、過去の情報を参照できなくなる。
- (3) 特徴空間補完で用いる特徴空間上で、過去の射影点履歴を検索する。(2)の検索空間である交通情報データ空間が欠損を伴うのに対して、特徴空間の射影点は後述のように観測データの相関成分を表すことから、欠損のない検索空間で履歴検索を行うことができる。よって、検索キーの減少によりヒット率が低下することはない。

以上の理由により、本研究で採用したのは(3)の手法である。本報告は開発の第一報であり、第2章でベース技術である特徴空間補完を概説した上で、第3章において予測手法のプロセスを解説する。第4章では本予測手法の妥当性の検証のため、プローブカーデ

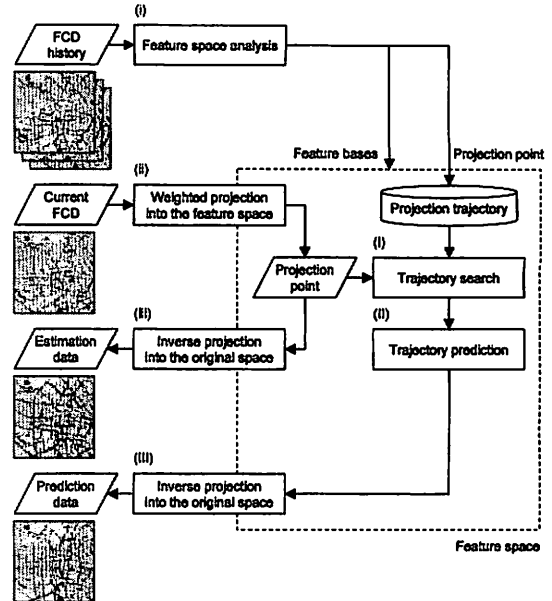


Fig. 1. Imputation and prediction process of floating car data.

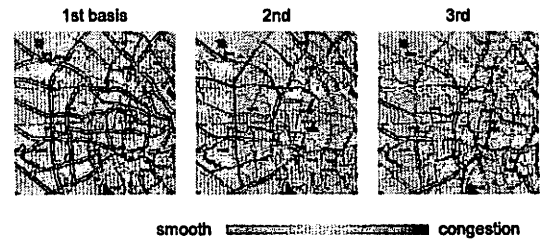


Fig. 2. Example of bases.

ータを用いた評価の前段階として、欠損のない路上センサデータを用いた予測結果について述べる。

2. 特徴空間補完

2.1. 基本概念

ある単位エリア(地図メッシュなど)で収集されたプローブカーデータを、リンク単位の旅行時間データに加工した上で主成分分析を行う。すると、複数リンクのデータが、相関をもって変化する成分と、無相関に変化する成分とに分解され、なおかつ、相関のある成分ごとに単一の代表変量で表されるため、データの次数が縮退される。たとえば、100リンクの旅行時間データを、10変量で代表することが可能である。

このとき、本来の旅行時間データは、前記代表変量を係数として、リンク間の相関関係を表す基準パターン(これを基底と呼ぶ)を線形合成することにより、近似的に表される。このように集約された情報表現が

特徴空間射影である。基底は特徴空間を張る静的なベクトルであり、前記代表変量が、特徴空間上で動的に変化する座標である。

さて、旅行時間データを特徴空間に射影して表現することが可能である以上、逆もまた可能である。すなわち、特徴空間上の一点を定め、それを逆射影すれば、当該単位エリアにおける交通状況を示すことになる。特徴空間に射影されるデータが欠損を伴う場合でも、特徴空間座標が定まるならば、逆射影されたデータはエリア全体の交通状況を表す。これが特徴空間補完の基本となる考え方であり、Fig. 1のように、

- (i) 過去のデータから特徴空間を生成し、
- (ii) リアルタイムに観測されたデータから、特徴空間上の座標を定め、
- (iii) 特徴空間座標の逆射影によって、推定情報を生成する、

という3つのプロセスで欠損データを補完する。

Fig. 2は東京都心の旅行時間データから得られた基底の事例である。主成分分析は基底に明示的な意味を与えないが、第1基底から順に都心全域の渋滞、都心外縁の渋滞、放射状路線の渋滞と解釈することができる。このように特徴空間が表すのは道路ネットワーク上の交通状況の相関関係であり、特徴空間を構成する基底ベクトルは、過去のデータから統計的に決定される。見方を変えれば、リンク間の相関関係は過去データから決定される静的な統計値ではあるが、それが多次元の情報であるがゆえに、その組み合わせ、すなわち特徴空間上の座標を動的に決定することにより、多様な交通状況の表現と推定補完が可能になる。

2.2. 推定補完プロセス

(i) 特徴空間の生成

路上センサで得られる欠損の少ないデータの場合には、特徴空間は通常の主成分分析によって生成できる。一方、プローブデータのように大規模な欠損を含む場合には、通常の主成分分析では演算処理そのものが成立しない。そこで、このようなデータに対しては、「欠損値付き主成分分析 (PCAMD)」^{[5][6]}と呼ばれる拡張手法を用いて、過去データからの特徴空間生成を行う。

具体的には、補完対象のエリアにおける M 本のリンクについて、 N 回にわたって計測された交通情報データを $N \times M$ 行列 X で表すものとする、 X の i 行目の成分を対角要素とするデータ行列 D_{xi} 、重み行列 V 、 V_0 に対して、PCAMD はフロベニウスノルム

$$J = \sum_{i=1}^N SS(D_{xi}V + V_0 - e_m u_i^T)_{D_{xi}} \dots \dots \dots (1)$$

を最小化する問題と定義できる。この問題を解くことで、処理対象の交通情報データ X の観測値を、誤差ノ

ルム最小で近似できる基底のセットが得られる。すなわち、交通情報データ X を、PCAMD で得られた基底で張られる特徴空間に射影すれば、その逆射影によって与えられるデータは、元の交通情報データに対する誤差最小近似となる。

(ii) 現況データの射影

前節で得られた基底に対して欠損の無い現況データを射影する場合には、基底と現況データの内積によって、特徴空間座標は一意に決定される。一方、現況データが欠損を伴う場合には、内積による射影は不可能であり、重み付け射影と呼ばれる次式を用いる。

$$a = (P^T W^T W P)^{-1} P^T W^T W x^T \dots \dots \dots (2)$$

ここで、 P は PCAMD で得られた基底ベクトルを並べた行列であり、 W は重み付けの行列である。欠損を含む現況データ x に対して、射影点 a が得られる。重み付け射影は、観測データの重みを 1、欠損データの重みを 0 として扱うことで、欠損データのリンクを無視し、現況データが観測されたリンクについて、特徴空間上の射影点と、射影前のデータの誤差ノルムが最小化されるように、射影点を決定する。すなわち、重み付け射影によって得られる特徴空間座標は、観測データに対する誤差最小近似値である。

ただし、ここで注意を要する事柄がある。前節で得られた特徴空間の次数にもよるが、現況データの欠損率が過度に高い場合には、重み付け射影問題を解くことができず、これが推定補完処理の可否を決める理論的な限界となる。すなわち、特徴空間補完は無から有を作り出しているわけではなく、処理対象となる単位エリアにおいて相関のある成分の次数に基づいて、その次数を超える現況データが得られた場合に、当該エリアの交通状況が推定可能であるものとして、補完データを算出する。

(iii) 逆射影による推定

前節の重み付け射影によって得られた特徴空間座標 a を、次式により元のデータ空間へ逆射影する。

$$\hat{x} = a^T P^T \dots \dots \dots (3)$$

ここで、特徴空間座標 a は、欠損データに対する推定を含む情報である。なぜならば、特徴空間座標は観測データに対する誤差最小近似値であるが、同時に、特徴空間が内包する相関関係に基づいて、全リンクの交通状況を表している。欠損データについて言えば、これは観測データとの相関で定まる値である。このことは 2.1 節に述べたとおりであり、いわば、特徴空間の持つ相関情報が、観測データと欠損データを関連付けている。よって、その逆射影により、

- ・ 現況データの観測値と

・過去データから得られたリンク間の相関関係に基づく推定値が得られる。

3. 特徴空間軌跡による予測

3.1. 基本概念

本節では、先述の特徴空間補完のプロセスをベースとして、予測を行う原理について述べる。

Fig. 3は式(2)によって得られた特徴空間上の射影点の軌跡である。Fig. 2の基底で張られた特徴空間に1ヶ月分の旅行時間データを5分間隔に射影し、特徴空間の第1, 第2軸ならびに第3, 第4軸で構成された部分空間の座標をプロットした。垂直軸は時刻であり、0:00~23:55の時間変化に相当する。すなわち、1日の射影点軌跡は0:00の平面上からスタートし、交通状況に応じて特徴空間上の座標を変えながら23:55の平面で終了し、再び0:00の平面に戻る。Fig. 4はFig. 3の第1軸, 第2軸の空間を平面に投影した図である。時刻はおおよその目安だが、1日単位で周期的な軌跡を描いていることがわかる。

主成分分析を用いると、元の旅行時間データ空間におけるデータ分散の大きい順に特徴空間軸である基底が決定される。Fig. 4の例において特徴空間の意味を解釈すると、第1軸は道路ネットワーク全体の交通量を表し(正/負:交通量が少ない/多い)、第2軸は外縁から都心への流入出に相当する(正/負:流出/流入)。すなわち、Fig. 4の射影点の変化を第1, 第2軸について定性的に説明すると下記ようになる。

- ・午前中は全体の交通量が増え、かつ、流入交通量が卓越するため、第1軸, 第2軸ともに負の方向へと射影点が移動する。
- ・昼には一旦交通量が減り、流入出が平衡するため、射影点は第1軸上で原点に近づく。
- ・午後には再び交通量が増え、かつ、夕方にかけて流出交通量が卓越するため、第1軸で負の方向に、第2軸で正の方向に射影点が移動する。

なお、射影するデータの欠損の有無に応じて軌跡の差異が生じるが、その違いは局所的な軌跡の振動であって、マクロな軌跡形状はよく似ている。

さて、リンク単体では同じ交通状況が1日に何度も発生するのは珍しいことではないが、これはリンク単体の旅行時間が、1次元空間上の値の増減にすぎないためである。一方、エリアとして交通状況をとらえ、特徴空間に射影することで、上述のように射影点の変化は1日単位の周期的な軌跡を描く。すなわち、特徴空間上のある領域に射影点が存在する場合には、その後の軌跡は過去の事例から類推可能であり、かつ、特徴空間上の座標が実際の交通状況と対応していることから、特徴空間上の軌跡の変化が予測できるならば、交

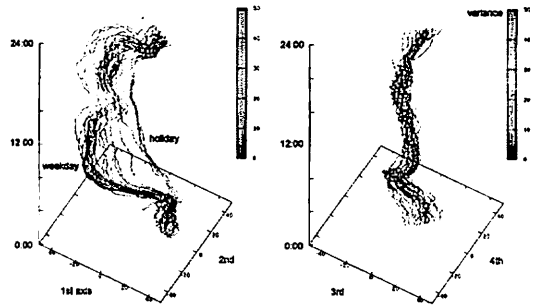


Fig. 3. Projection trajectory in the feature space.

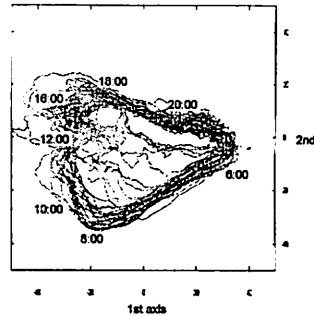


Fig. 4. Cyclic shape of the feature space trajectory.

通状況の変化も予測できる。これが特徴空間軌跡を用いた予測の基本概念であり、Fig. 1における推定補完の処理(i)(ii)(iii)と並行して、

- (I) 現況データの射影点近傍にある過去データの射影点履歴を検索
- (II) 検索された射影点を起点とする軌跡に基づいて未来の射影点を予測
- (III) 予測射影点を旅行時間データ空間に逆射影というプロセスにより、現況データを基にして未来の交通状況を予測する。次節においては、このプロセスに沿った予測の詳細を述べる。

3.2. 予測プロセス

(I) 射影点の検索

Fig. 5は予測の原理を示した図である。まず、式(2)によって特徴空間に射影された現況データの射影点について、Fig. 4のような過去の射影点の軌跡から、その近傍にある複数の射影点を検索する。具体的には、現在時刻 t_0 における射影点 $a(t_0)$ に対して、過去の時刻 t における射影点 $a(t)$ の距離

$$d(t) = (a(t) - a(t_0))^{1/2} \dots \dots \dots (4)$$

を計算し、この結果に基づいて過去一定期間の射影点履歴をソートし、その上位 K 個の射影点 $a(t_1) \sim a(t_K)$ を以下の予測演算に用いる。

(II) 予測射影点の算出

現在時刻 t_0 から Δt 後の予測を行う場合には、(I) で得られた K 個の射影点 $a(t_1) \sim a(t_K)$ を起点として射影点軌跡を辿り、各射影点の Δt 後の射影点 $a(t_1+\Delta t) \sim a(t_K+\Delta t)$ を得る。特徴空間上で現況の射影点 $a(t_0)$ との距離が小さい、すなわち、現在の交通状況と類似の状況を示す過去の射影点ほど、その後の変化も類似していると考えられるならば、射影点 $a(t_1+\Delta t) \sim a(t_K+\Delta t)$ を距離 $d(t)$ で重み付けした重心

$$\hat{a}(t_0 + \Delta t) = \sum_{i=1}^K \frac{1}{d(t_i)^2} a(t_i + \Delta t) \dots\dots\dots (5)$$

が、過去の射影点履歴から類推しうる範囲で、 Δt 後の射影点として最も確かな予測値である。予測情報として静的な統計データを提供する手法と対比して考えると、式 (5) の処理は、現在の交通状況に基づいて動的に統計処理対象の母集団を定め、現況に応じた統計的代表値を生成する処理とも言える。

(III) 予測射影点の逆射影

式 (5) で得られた予測射影点を、式 (3) と同様の演算によって旅行時間データ空間に逆射影する。すなわち、

$$\hat{x}(t_0 + \Delta t) = \hat{a}^T(t_0 + \Delta t) P^T \dots\dots\dots (6)$$

が旅行時間データの予測値である。また、射影点座標を用いて基底ベクトルを合成することから、この式は欠損データの推定値を算出する作用も示す。それゆえ、欠損を伴うプローブカーデータに基づく予測が可能となる。

4. 予測精度の評価

4.1. 評価手順

緒言で述べたように、プローブカーデータによる評価の前段階として、予測手法の妥当性を検証することを目的として、欠損のない路上センサデータを用いた予測結果について述べる。評価手順を以下に示す。

- (1) 過去の交通情報から、特徴空間を生成する。
使用したデータは、東京都心部 (Fig. 6 の点線領域) の 400 本のリンクにおいて収集された 2004 年 6 月 1 日～28 日の旅行時間データである。この処理は 2.2 節 (i) に相当する。
- (2) (1) のデータを特徴空間に射影し、3.2 節 (I) において検索対象となる射影点履歴を生成する。データが 5 分周期であることから 1 日当たりの射影点は 288 サンプルであり、当該期間を通して、8064 サンプルの射影点履歴が検索空間となる。
- (3) 2004 年 6 月 29 日～7 月 26 日を評価期間とし

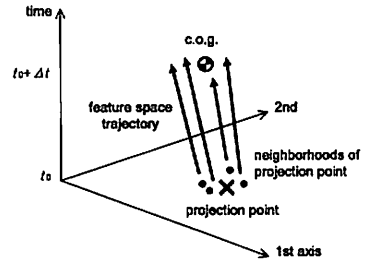


Fig. 5. Basic idea of feature space prediction.

て、その期間の旅行時間データを現況データとして特徴空間に射影する。この処理は 2.2 節 (ii) に相当する。

- (4) (3) で得られた各射影点ごとに、3.2 節 (I) によって、(2) で生成した射影点履歴から近傍射影点の検索を行い、(II) によって予測射影点を計算し、(III) によって旅行時間データの予測値を算出する。ここでは、2 時間先の予測を行うものとして、式 (5) (6) において $\Delta t = 120$ [分] とした。また、全射影点履歴 8064 サンプルから、 $K=100$ として上位 100 サンプルを式 (5) の重心演算に用いた。これは、射影点履歴のうち、1 日あたり 2～3 サンプル程度が現況データに類似するであろうと想定して設定した数値である。もちろん、上位 100 サンプルに入ったとしても、現況射影点との距離が大きい射影点は、重心演算における寄与率は低くなる。
- (5) (4) で得られた予測値を、(3) の旅行時間データにおける 2 時間先の真値と比較して、予測誤差を算出する。

4.2. 評価結果

以上の手順によって得られた予測結果の一例を示す。Fig. 7 は 1 リンクの旅行時間データの予測値と実測値を比較したグラフであり、横軸が日時、縦軸がリンク旅行時間である。4.1 節の (2) で生成した検索空間は曜日で分類されたものではないが、Fig. 7 においては、実測値に見られる曜日による交通状況の違いが、予測値にもよく現れている。これは、3.2 節で述べた予測プロセスが日時に依存することなく、現況値として得られたデータに基づいて予測値を算出することによる。曜日、祝日、季節などの日種分類は一般に人手による作業を介在するため、日種分類を必要としないことは、交通情報システムの運用に際して非常に大きなメリットとなる。

また、Fig. 7 のグラフにおいては、渋滞の立ち上がり、立ち下がりにおける遅延がほとんど見られない。一般に、現況データを入力として過去の事例を参照す

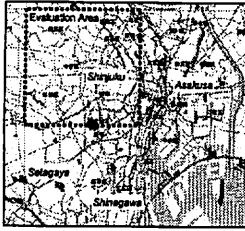


Fig. 6. Evaluation area.

る事例ベースの予測手法においては、現況データの立ち上がりを捉えてから予測処理が追従するため、予測結果の遅延は避けられない。これは本予測手法においても同様のはずだが、各リンク単位では微かな変化であっても、それらが相関をもって変化する場合には特徴空間上での顕著な変化として現れることから、渋滞の発生を即座に捉えることが可能になり、遅延が解消されたのではないかと考えられる。このことは本予測手法の優位性を示す上で、今後、理論的に検証していきたいテーマである。

Fig. 8は評価対象とした400リンク全てについて、実測値に対する予測値の誤差分布を示したヒストグラムである。直線で囲まれた領域が予測誤差率 ± 0.3 の範囲であり、予測結果の83%がこの範囲に分布することから、精度良く予測できていることがわかる。評価期間を通して、予測誤差率の絶対値の平均は0.16であった。なお、平均的なリンク長は約600mである。

以上は欠損のない路上センサデータによる評価結果であって、本来の目的であるプローブカーデータの利用については、今後評価を行っていく。予測対象データが欠損を伴う場合には、Fig. 3の特徴空間軌跡に局所的な振動が生じるが、式(5)に示した重心演算は、射影点の振動に対してもロバストな予測値を得ることを目的としたものである。プローブカーデータを対象とした評価においては、このような予測結果のロバスト性を中心に検証を進める計画である。

5. 結言

本研究では、プローブカーデータに基づく予測情報の提供を目的として、特徴空間射影を用いた予測手法の開発を行った。これは、同様に特徴空間射影を用いたリアルタイム推定補完手法をベースとする技術であり、欠損を伴うプローブカーデータから、欠損のない予測情報の演算を行う技術である。現況のプローブカーデータを入力として特徴空間上で過去の事例を参照するため、事前に曜日や季節による事例分類を行う必要はなく、実サービスにおける運用性の高さが期待できる。同時に、事例ベースの予測技術に付き物である渋滞の立ち上がり／立ち下がりにおける遅延がほとん

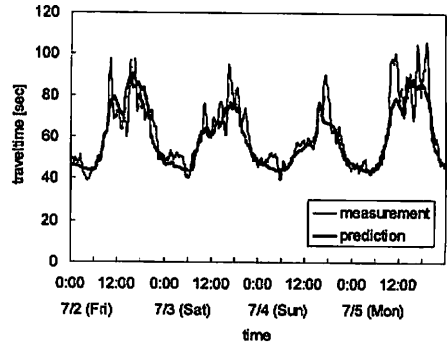


Fig. 7. Example of prediction.

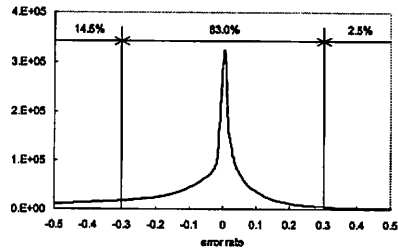


Fig. 8. Distribution of prediction error.

ど生じないことも優位性として確認された。

今回の検証は欠損の伴わない路上センサデータを対象とした結果だが、今後はプローブカーデータを用いた検証を進めていく計画であり、その過程においては、予測結果のロバスト性を重視した技術開発を進めていきたいと考えている。

文 献

- [1] 柘植正邦, 養父利秀, “カーナビゲーションの可能性を広げる新情報提供システム,” 自動車技術, vol.58, no.2, pp.44-48, Feb.2004.
- [2] W. Wang, T. Fushiki, H. Zhang, K. Kimita, and H. Mizuta, “A Probe Car System for Beijing,” Proc. 12th World Congress on ITS, CD-ROM, San Francisco, Nov.2005.
- [3] 伏木匠, 横田孝義, 君田和也, 熊谷正俊, “交通情報提供の更新周期を考慮したプローブカー台数算出モデルの定式化とその評価,” 電気学会論文誌 D, vol.126-D, no.6, pp.741-747, Jun.2006.
- [4] 熊谷正俊, 伏木匠, 横田孝義, 君田和也, “特徴空間射影によるプローブカーデータのリアルタイム補完,” 情報処理学会論文誌, vol.47, no.7, pp.2133-2140, Jul.2006.
- [5] A. Ruhe, “Numerical computation of principal components when several observations are missing,” Tech. Rep. UMINF-48, Dept. Information Processing, Umea Univ., 1974.
- [6] 高根芳雄, 制約付き主成分分析法, 朝倉書店, 1995.