

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

熊谷正俊[†] 伏木 匠[†]
横田孝義[†] 君田和也[†]

プローブカーデータの欠損をリアルタイムに補完し、欠損のない交通情報を提供する技術について述べる。プローブカーは路上センサに依存せずに交通情報を収集するうえで有効な手段である。しかしながら、一般にプローブカーの走行経路が不確定であることから、プローブカーデータの時系列には大規模な欠損が生じる。それゆえ、プローブカーシステムの導入コストを抑えつつ、欠損のない交通情報を提供するには、プローブカーデータの特性に合わせた補完技術が必要となる。補完手段として最も簡単なのは、過去のデータ履歴の同時刻平均値を補完データとすることである。しかし、この手法では日々異なる交通状況の変化を反映できない。また、回帰分析はデータ補完の一般的な手法だが、欠損率が高い場合には十分に機能しない。欠損率が高いプローブカーデータについても高精度な補完を実現するために、本論文では特徴空間射影に基づく新たな動的補完手法を提案する。この補完手法は大きく分けて、(1)過去のプローブカーデータ履歴に基づく特徴空間の形成、(2)欠損を含む現況データの特徴空間への射影、そして、(3)特徴空間からの逆射影による欠損値の補完という3つのプロセスから構成されている。特徴空間を介して処理を行うことで、特徴空間の主軸、すなわち複数リンク間で相関のある成分ごとに欠損情報の推定がなされるため、観測値との相関に基づく動的かつ高精度なデータ補完が可能となる。

Realtime Interpolation Method for Probe Car System Based on Feature Space Projection

MASATOSHI KUMAGAI,[†] TAKUMI FUSHIKI,[†] TAKAYOSHI YOKOTA[†]
and KAZUYA KIMITA[†]

This paper discusses a realtime interpolation method for a probe car system which provides traffic information with no missing data. Probe cars are an effective way to collect traffic information without roadside sensors. However, because of the random routes of probe cars, there is a large amount of missing data in the time series of probe car data. Therefore, in order to minimize the cost of a probe car system and provide continuous information, a data interpolation method appropriate for probe car data is required. The easiest means of interpolation is to provide a time-sliced average of the probe data history. However, the time-sliced average cannot describe traffic conditions that change day by day. Though regression analysis is a suitable method for data interpolation, it does not work well when the rate of missing data is too high. In order to perform an accurate interpolation with probe car data, we propose a new statistical method based on feature space projection. The method consists of three major processes: (1) determination of a feature space from past probe data history; (2) feature space projection of current probe data; and (3) interpolation of missing data performed by inverse projection from the feature space. Since interpolation of the missing information is achieved on each feature space axis that represents a correlated component of the traffic data, the method can accurately estimate the missing information based on the correlation with the observed information.

1. 緒 言

交通情報サービスの役割は、渋滞箇所をドライバに知らせ、最適な経路へと導き、また、所要時間を見積もることである。社会的な観点からは、限られた道路

資源を有効に活用し、需給バランスを改善することで、環境負荷の低減、渋滞による経済損失の抑制などの効果が期待されている。

当初の交通情報サービスは、特定の道路区間の現況情報を提供することからスタートしたが、今日ではさまざまな形に多様化をとげつつあり、その方向性は大きく分けて2つある。1つは、予測情報の提供^{1)~4)}という時間方向の拡張である。予測情報の利用により、

[†] 株式会社日立製作所日立研究所
Hitachi Research Laboratory, Hitachi Ltd.

長距離の移動や長期的な配送計画においても、適切な経路選択や所要時間の見積りが可能になる。そしてもう一つは、プローブカーによる交通情報提供エリアの拡大^{5),6)}、すなわち空間方向の拡張である。既存の路上センサではカバーされていない道路についても交通情報の収集・提供を行うことで、より効果的なサービスの確立を目指す。

本論文のテーマは後者の課題、プローブカーによる情報提供に関するものである。プローブカーシステムでは車両自体がセンサであるため、路上インフラに依存しない交通情報の収集が可能であり、原理的にはすべての道路における交通情報の提供が可能である。しかしながら、センサである車両の走行位置が一般に不確定であることから、その情報品質は路上センサで収集される連続的なデータとは大きく異なり、空間的・時間的に大きな欠損を含む。

たとえば、プローブカーの台数を全国で10万台とした場合、プローブカーデータが取得できる時間密度は、道路リンクあたり1時間に1,2回程度である⁷⁾。これを現行の路上センサと同等の5分周期のデータとして利用するうえで、時系列上での欠損率は8割から9割に達する。このように疎らなプローブカーデータを路上センサと相補的な情報源として活用するには、何らかのデータ補完手段が必要となる。

補完手段として最も一般的なのは、過去に計測されたプローブカーデータから同時刻平均値などの代表値を算出しておき、現況のプローブカーデータが欠損した場合には、それを補完情報として提供する手法である。この手法は簡単かつ安定した補完情報の提供が可能だが、日々異なる交通状況に対して十分に追従できないという問題がある。この問題は所要時間の見積り誤差として現れるとともに、実際のサービスシーンにおいては、現況の交通状況に関係なくいつも同じ場所が混んでいる、あるいは同じ経路誘導が行われるという形で情報提供が行われることになり、サービスの魅力を半減させる。正確かつ魅力的な交通情報サービスの提供には、現況の交通情報に立脚した補完処理が不可欠といえる。

本論文では以上の観点から、新たに開発したプローブカーデータのリアルタイム補完手法を提案する。この手法は、過去のプローブカーデータから特徴空間を生成し、現況のプローブカーデータをその欠損に応じて特徴空間に射影することで、欠損値の補完を行う。特徴空間射影により補完を行うことの利点は、複数の道路リンクの間で相関のある成分に交通情報を分解し、その成分ごとに補完データの算出および合成を行うこ

とで、過去データに含まれる多様な相関情報を活用した補完処理が可能となることである。

以下は本論文の構成である。2章では新たに開発した特徴空間データ補完の基本アルゴリズムについて、また、3章では補完精度の向上に有効な拡張手法について述べる。4章はその効果を評価した結果である。5章は結言であり、今後の課題、展望などについて述べる。

2. 特徴空間データ補完

本手法の基本的な考え方を模式的に表すと図1(1)のようになる。図1(1)において、等号の左辺は複数の道路リンクにおけるある瞬間の交通情報(旅行時間など)の値を線の太さで表したものであり、右辺はそれを複数の基底の線形合成として表記したものである。右辺において、基底のそれぞれは各リンクにおいて相関を持って変化する交通情報の成分で構成され、各基底の係数は無相関に変化する。交通情報をこのように表現することで、複数リンクにおける交通状況の傾向を、各基底の係数の大きさによって表すことができる。

たとえば、基底1におけるリンク1, リンク2, リンク3それぞれの成分を

$$[l_{11}, l_{12}, l_{13}] = [0.1, 0.1, 0.5]$$

とすると、それはリンク1~3の交通情報に“1:1:5”という比例関係で変化する成分が含まれていることを意味する。一方で基底2において、リンク1~3それぞれの成分が

$$[l_{21}, l_{22}, l_{23}] = [0.3, -0.3, 0.0]$$

であれば、先の“1:1:5”という比例関係とは無相関に、“3:-3:0”という比例関係で変化する成分も含まれていることになる。そして、“1:1:5”で変化する成分の強度(基底1の係数 a_1)と、“3:-3:0”で変化する成分の強度(基底2の係数 a_2)によって、

- リンク1とリンク2に比べて、リンク3が卓越して渋滞している

● リンク1が渋滞して、リンク2は空いている
など、リンク1~3の交通状況がいかなる傾向にあるか、表現することができる。

このようなリンク間の相関関係を表す基底がいくつ存在するかは、当該リンク群の交通情報が持つ成分の次元によって決まる。その決め方は、信号の成分分解の代表的な手法であるフーリエ級数と対比して考えると分かりやすい。フーリエ級数において、任意の周期関数は複数の余弦波および正弦波の線形結合で表され、次数が上がるごとに周波数は2倍になり、それらは互いに直交する基底である。ここでは処理対象の情報が有限長であり、また、扱う情報に応じて固有の基底を

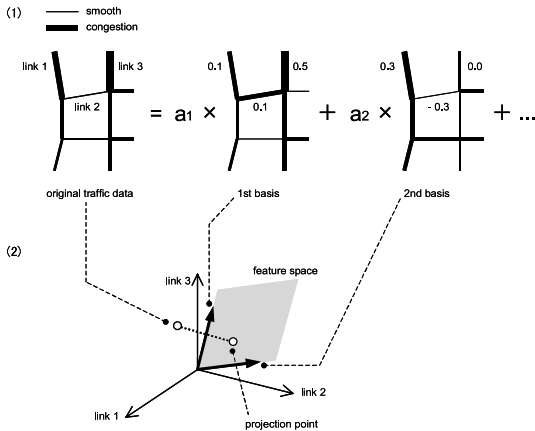


図 1 複数道路リンクにおける交通状況の基底による表現
Fig. 1 Traffic condition of multi road-links represented by feature bases.

用意するという差異はあるが、基底の数が情報の次元で定まるのはフーリエ級数と同様である。実用に際して、交通情報の表現に十分な基底の数は、各基底に対応する成分の強度、すなわちフーリエ級数でいうところのスペクトルを見ることで決定できる。

図 1 (1) の考え方を特徴空間という観点でとらえると、図 1 (2) のように、基底のそれぞれは特徴空間を構成する主軸ベクトルであり、その係数は特徴空間座標に相当する。すなわち、当該リンク群におけるある瞬間の交通状況は、特徴空間内の 1 点によって近似的に表される。

逆に、現況の交通情報がプローブカーデータのように大きな欠損を含むものであっても、それを特徴空間に射影することができれば、射影された特徴空間座標を元の交通情報データ空間に逆射影することによって、交通情報の欠損したリンクを推定補完することができる。このような特徴空間射影を用いた推定補完には、複数リンクの間で相関のある成分ごとに推定値が算出されるので、さまざまな要因による複合的な現象として観測される交通情報を適切に補完できるというメリットがある。

以下、本手法の具体的なプロセスを図 2 のステップ 1~3 に沿って解説する。

ステップ 1

先述のような基底を主軸ベクトルとする特徴空間は、主成分分析によって得ることができる。ただし、プローブカーデータのように欠損率が高いデータを扱う場合には、通常のパワイズ、リストワイズな欠損値除去手法を用いることはできない。相関関係の解析に必要なサンプルが、欠損値除去によって皆無となるためである。

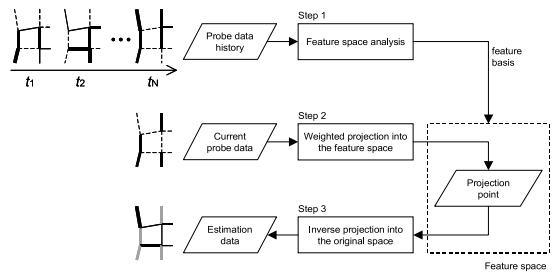


図 2 プローブカーデータの補完プロセス
Fig. 2 Interpolation process of probe car data.

そこで本研究では、“欠損値付き主成分分析 (PCAMD)” と呼ばれる主成分分析の拡張手法を用いて、特徴空間の生成を行った。PCAMD はサンプルデータのうち、観測されたデータのみを用いて特徴空間を決定する手法であり、Ruhe によってそもそも提案された数値的な手法⁸⁾と、後年、柴山らによる欠損値を加味した線形等化法⁹⁾を高根が多次元に拡張した解析的な手法¹⁰⁾がある。ここでは大規模なデータ解析における安定した処理を重視し、後者の解析的な手法を用いることにした。

PCAMD による特徴空間生成のプロセスを下記に示す。補完処理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。

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 からなる

$$Y_i = D_{xi}V + V_0 \tag{1}$$

に対して、PCAMD による特徴空間基底の決定は、制約条件

$$V^T S V = I \tag{2}$$

の下で、SS ノルム¹⁰⁾

$$J = \sum_{i=1}^N SS (Y_i - e_M u_i^T)_{D_{wi}, I} \tag{3}$$

を最小化するという問題である。ここで、 S は観測値に関する各リンクのデータ変動を対角要素とする行列、 D_{wi} は D_{xi} の対角要素が観測値の場合は 1、欠損値の場合は 0 となる行列である。この解は、

$$Q_{M/D_{wi}} = I - e_M e_M^T D_{wi} / \text{tr}(D_{wi}) \tag{4}$$

$$C_i = Q_{M/D_{wi}}^T D_{wi} Q_{M/D_{wi}} \tag{5}$$

$$A_1 = \sum_i D_{xi} C_i D_{xi} \tag{6}$$

$$A_2 = \sum_i D_{xi} C_i \tag{7}$$

$$A_3 = \sum_i C_i \tag{8}$$

$$A = A_1 - A_2 A_3^+ A_2^T \quad (9)$$

に対して、固有値問題

$$S^{-1/2} A S^{-1/2} \tilde{V} = \tilde{V} \Delta \quad (10)$$

を解くことによって得られる。ただし、

$$V = S^{-1/2} \tilde{V} \quad (11)$$

$$V_0 = -A_3^+ A_2^T V \quad (12)$$

$$u_i = e_M^T D_{wi} Y_i / \text{tr}(D_{wi}) \quad (13)$$

である。ここで得られる解のうち、小さな固有値に対応する固有ベクトルを並べたものが特徴空間基底行列 P である。

ステップ 2

過去の交通情報から得られた特徴空間基底行列 P に対して、現況のプロブカーデータ x を射影し、特徴空間座標を決定する。ただし、現況データ x に欠損があることから、内積による射影を行うことはできない。特徴空間は元の交通情報データの部分空間であることから、交通情報データ空間の座標軸に沿った重み付けを行い、特徴空間座標は現況データ x に対する誤差ノルム最小解として、次式により得られる。

$$a = (P^T W P)^{-1} P^T W x^T \quad (14)$$

ここに W は重み行列であり、現況データ x の観測値に対しては 1、欠損値に対しては 0 とする。

このような重み付けをすることで、射影点は現況データの観測値のみによって決定される。一方で、特徴空間はその主軸ベクトル自体がリンク間の相関関係を表す情報である。すなわち、式 (14) による特徴空間座標は、現況データの観測値を反映し、かつ、欠損値に対する推定情報も内包することになる。

ステップ 3

式 (14) で得られた特徴空間座標を、次式により元の交通情報データ空間に逆射影する。

$$\hat{x} = a P^T \quad (15)$$

逆射影で得られた \hat{x} は、特徴空間上の射影点が x に対する誤差ノルム最小解であるという性質から、 x の観測値に対してはその近似値であり、また、特徴空間がリンク間の相関関係を表すことから、 x の欠損値に対する推定値である。 x の欠損値を \hat{x} で置き換えることで、 x の補完がなされる。

3. 精度向上のための時間的・空間的拡張

3.1 現況データのタイムウィンドウ拡大

前章に述べた手法を用いれば、過去の交通情報の相関に基づいて、現況データを反映した欠損値の補完が可能になる。とはいえ、現況データの欠損率が高くなるほど、補完精度が低下することは否めない。情報量

が少なく、欠損パターンがランダムに変化する交通情報から特徴空間射影を行うことで、特徴空間座標がわずかな観測値に大きく左右されるためである。

また、主成分分析本来の性質から、特徴空間の次数が高いほど元の交通情報データに対する表現力が高くなり、変化の小さなリンクも補完できるようになるが、一方では特徴空間の次数を上げると、寄与率の低い基底の影響により、補完精度がかえって悪化するという問題がある。下位の基底は、事故や工事による渋滞などのマイナーな情報を含む傾向にあり、そのような基底で張られた特徴空間に欠損率の高い現況データを射影した場合、欠損パターンによっては、それら過去の突発事象に起因した成分が過度に現れるためである。

より欠損率の高いデータに対応し、また、特徴空間の次数を上げて交通情報の再現性を高めるには、単純に考えれば、前章ステップ 2 において、特徴空間に射影する現況データのタイムウィンドウを長くとり、ある程度過去にさかのぼって観測値を多めに確保すればよい。しかしながら、たとえば朝夕のラッシュ時間帯などには、わずかな時間でも交通状況が大きく変化することから、20 分程度前の比較的新しい情報を射影に用いたとしても、補完情報には顕著な遅延が生じる。

そこで本研究では、観測値の確保と遅延の回避を両立させるため、式 (14) の重み行列 W をデータのタイムスタンプに応じた可変値として、重み付け射影問題を解くことを考えた。たとえば、現在時刻におけるタイムスタンプを 0 として、データ更新周期の長さだけ過去にさかのぼるごとに、タイムスタンプを 1 増やすものとする。式 (14) の定義と同様に、重み行列を 1 (観測値) と 0 (欠損値) で構成し、最新のタイムスタンプに対する重み行列を W_0 、タイムスタンプ i に対する重み行列を W_i とすると、式 (14) における重み行列 W は

$$W = \sum_{i=0} z_i W_i \quad (16)$$

である。ここに、 z_i はタイムスタンプに応じた係数であり、 $z_0 = 1, 1 > z_1 > z_2 > \dots > 0$ と設定することで、古いデータによる遅延の影響を緩和しつつ、特徴空間座標を安定に決定するための観測データをより多く確保することが可能になる。

3.2 路上センサデータの併用

前節が特徴空間に射影するデータの時間的拡大であるのに対し、本節ではその空間的な拡大として、路上センサデータとプロブカーデータの併用による補完精度向上について述べる。路上センサはインフラ整備に多大なコストがかかるため、そのカバーエリアは限

られているが、情報としては常時収集可能な安定したデータであり、プローブカーデータと合わせて特徴空間の生成ならびに射影による補完処理を実施することで、補完データの安定化と精度向上が可能になる。

2章ステップ1の特徴空間生成において、プローブカーデータと路上センサデータの利用に大きな区別はない。プローブカーデータ行列 X_P と路上センサデータ行列 X_R を部分行列とする

$$X = [X_P \ X_R] \quad (17)$$

を交通情報データ行列として式(4)~(10)を解き、特徴空間を構成する有意な情報量を増やし、特徴空間座標の決定を安定化する。このとき、路上センサに対応する D_{wi} の対角要素をすべて1とすることで、路上センサで連続的にデータが収集されるリンクと、プローブカーで疎らにデータが収集されるリンクとを同一の仕組みで取り扱うことができる。

一方、ステップ2の特徴空間射影においては、アルゴリズムの拡張が必要となる。路上センサとプローブカーのカバーエリア比率や、プローブカーデータの欠損率にもよるが、同時刻に路上センサデータの計測されるリンクの数が、プローブカーデータのそれに比べて多ければ、推定補完の結果に対して路上センサの情報が支配的になる。また、実際の交通情報システムにおいて、路上センサは主要道路の連続したリンクを中心に設置されており、そのような区間ではある特定の相関情報が現れる傾向にある。すなわち、道路ネットワーク全体の推定処理が路上センサ設置区間の交通状況に過度に依存することになり、プローブカーでリアルタイムに情報を収集し、そのデータを元に推定補完を行う効果が薄れてしまう。

そこで、ここでは式(14)の重み行列 W をデータ種別に応じた可変値として、この問題に対処する。重み行列 W は特徴空間射影における残差ノルムの最小化に際して、座標軸ごとの尺度を変える作用を持つ。すなわち重みの大小によって、射影点の決定における各リンクの寄与の度合いを調節することができる。それゆえ、プローブカーデータの重みに対して、路上センサデータの重みを低く設定することで、推定補完処理において路上センサデータを参照しつつ、その結果に対する依存度を適度に緩和することが可能となる。

この考え方に基づいて重み行列 W を表記すると、次式のようになる。

$$W = \begin{bmatrix} z_P W_P & 0 \\ 0 & z_R W_R \end{bmatrix} \quad (18)$$

ここに、 W_P 、 W_R はそれぞれプローブカーデータと

路上センサに対する部分重み行列であり、 z_P 、 z_R は両者の係数である。また、以上のように0, 1以外の重み付けを行う場合には、式(14)を

$$a = (P^T W^T W P)^{-1} P^T W^T W x^T \quad (19)$$

という形で解く。

4. 補完精度の評価

補完精度の評価は図3に示した20km四方の領域における180リンクについて行った。ただし、現時点では広域の評価に必要なプローブカーデータが収集の途上にあるため、路上センサのリンク旅行時間データをランダムに欠損させ、模擬プローブカーデータとして評価に用いた。プローブカーデータを用いた評価結果については、路上での実験を経て後日報告したい。評価手順は下記のとおりである。

- (1) 2001年4月1日~7日の模擬プローブカーデータおよび路上センサデータから、2章ステップ1で述べた特徴空間生成を行う。
- (2) (1)で得られた特徴空間に対して、2001年4月8日~14日の模擬プローブカーデータおよび路上センサデータを2章ステップ2に従って射影し、データ更新周期ごとの特徴空間座標を決定する。
- (3) (2)の特徴空間座標を元のデータ空間に逆射影し、データ更新周期ごとの補完値を算出する。

以上のプロセスによる補完精度の評価は、(2)の期間において模擬プローブカーデータに欠損を与える前の真値と、逆射影によって得られた補完値との比較によって行う。

表1は真値に対する補完値の誤差パーセントについて、すべての欠損リンクから時刻ごとに平方二乗平均(RMS)を計算し、さらに(2)の期間を通して時刻ごとRMS誤差を平均した値の表である。ここでは同時刻平均値を用いた従来手法(平日/休日分類)と提案手法について評価を行った。提案手法に関する評価パラメータは下記のとおりである。

- 特徴空間次元 D : 5次元, 10次元
- 特徴空間射影に用いる現況データの時間幅 T : 0分, 20分
(データ更新周期は10分であり、 T を0分とする場合には式(16)におけるタイムスタンプは $i=0$ のみ、 T を20分とした場合には $i=0, 1, 2$ である)
- 路上センサデータ: あり, なし

従来手法における模擬プローブカーデータの欠損率は80%として、特徴空間データ補完については80%と



図 3 評価エリア
Fig. 3 Evaluation area.

表 1 補完誤差の比較
Table 1 Comparison of interpolation error.

	traditional	80% missing				90% missing			
		D=5		D=10		D=5		D=10	
		T=0	T=20	T=0	T=20	T=0	T=20	T=0	T=20
probe data only	38.3	28.7	26.0	31.2	26.8	33.6	29.9	/	/
road sensor combined		25.7	24.9	27.4	25.6	28.4	27.6	31.2	28.8

90%の場合の比較も行った。表中、斜線の箇所は、データの持つ情報量が所定の次数に足りず、特徴空間を生成できなかったケースである。

模擬プローブカーデータの対象として用いたリンクの数は、全 180 リンクのうち 150 リンクであり、それらにランダムな欠損を与える。残りの固定 30 リンクが路上センサデータに相当し、常時データが存在する。すなわち、欠損率 80%とした場合の評価データの内訳はたとえば図 4 のようなものである。ここで、表 1 における誤差評価の対象としたデータは、図 4 における欠損データ約 120 リンクのみであり、評価条件において路上センサデータありとした場合でも、模擬プローブカーデータの欠損率が下がるわけではなく、評価対象であるプローブカーデータの品質は同等である。また、路上センサデータとして用いたリンクと、模擬プローブカーデータとして用いたリンクは重複せず、隣接するリンクでもないので、路上センサデータのみを用いた補完はできない。

式 (16) に示したタイムスタンプに応じた重み付けの係数は、 $[z_0, z_1, z_2] = [1.0, 0.8, 0.6]$ とした。また、路上センサデータの併用におけるプローブカーデータと路上センサデータとの重み付けの係数は、 $[z_P, z_R] = [1.0, 0.3]$ とした。これらは実験的に決

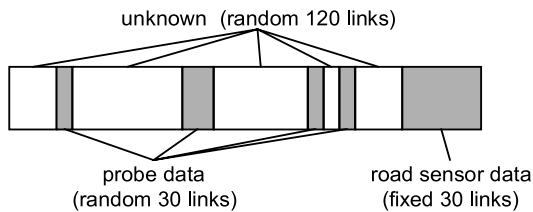


図 4 評価データの内訳
Fig. 4 Components of evaluation data.

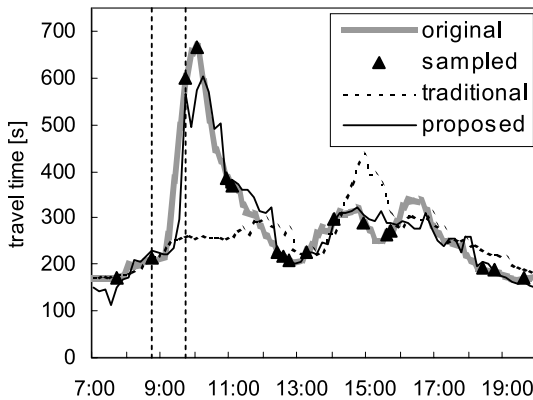


図 5 補完データの例
Fig. 5 Example of interpolation data.

定した値だが、今回の評価においては、前者はおおよそ ± 0.1 の範囲内で、また後者は ± 0.2 程度の範囲内で、補完精度にさしたる影響を与えないことが分かっている。重み付けの係数は式 (19) の誤差ノルム最小化において、その 2 乗で作用する項であり、偏微分すれば 1 次の項となる。ゆえに、重み付けの係数の絶対値が小さいほど、その変動が式 (19) による射影点 a の決定に及ぼす影響は小さく考えられるが、この問題に関する詳細の考察ならびに欠損率やリンク数に応じた決定方法は、今後の課題としたい。

表 1 の結果からは次のようなことが分かる。まず、従来手法に対して、特徴空間データ補完を用いることにより、補完精度の大幅な改善が可能である (RMS 誤差にして 10 ポイント)。たとえば図 5 はあるリンクについて、従来手法と提案手法の推定値を比較したグラフだが、同時刻平均値のような統計的代表値に現れない渋滞がある場合でも、提案手法を用いれば推定補完を行うことができる。また、特徴空間射影対象データのタイムウィンドウの拡大、ならびに路上センサデータの併用には次のような効果がある。

補完誤差低減の効果

補完誤差をさらに数ポイント低減させることができる。射影データのタイムウィンドウを広げる場合に

はおおむねその時間幅に比例して、また、路上センサーデータを併用する場合には、常時一定量の観測データが特徴空間射影に寄与し、射影点が安定に決定されるためである。朝の渋滞のように変化が急峻な時間帯においては、路上センサーデータ併用のみを適用し、より追従性を重視した補完処理を行うことも考えられる。

なお、交通情報の予測における誤差の数ポイントの違いというのは、単純にいえば、所要時間 100 分の予測に対して誤差が数分減るということである。それが大きな意味を持つかどうかは利用者によって意見の分かれるところだが、一方で、大半の利用者の関心事項は、目的地により早く着く経路情報の提供にある。いうまでもなく、適切な経路探索にはより正確な交通情報が不可欠であり、そのことが交通情報事業者をして、交通情報のカバーエリアを広げ、誤差を少しでも減らそうという研究開発の原動力になっている。

補完データの分解能向上の効果

プローブカーデータ欠損率が高い場合には、特徴空間次数を増やすことで補完精度は悪化し、特徴空間射影の演算が成立しないこともある。前者の問題は先述のとおり、下位の基底ほどマイナーな成分を含み、限られた観測データからその強度を安定に決定するのが困難になるためであり、後者は観測データのランク落ちに起因する。

一方で、主成分分析の性質上、特徴空間の次数が高いほどデータの再現性は高まるので、補完データの分解能向上には、下位の基底も必要である。すなわち、特徴空間次数を限定した場合には、メジャーな状況において補完データが不安定にならない分、補完精度は全体として向上するが、マイナーな状況への追従性は下がり、局所的に見れば誤差は増大するといえる。

この問題に対して、タイムウィンドウ拡大、路上センサーデータ併用のいずれも、特徴空間次数の拡大と補完誤差の低減を両立させる効果がある。特に両者を同時に適用した場合には、特徴空間次数を 2 倍にしても遜色ない補完精度が達成できた。

欠損率の高いデータを補完する効果

欠損率 80% と 90% の場合を比較すると、路上センサーデータ併用の効果は、プローブカーデータ欠損率が高いほど大きい。プローブカー導入の初期において、仮に日本全国の主要道路に 10 万台のプローブカーが稼働していたとしても、空間的なデータ欠損率は 90% に達するが⁷⁾、そのような段階でこそ補完データの提供は重要である。ゆえに、実用サービスを立ち上げるうえで、路上センサーのない路線で収集された限定的なプローブカーデータをもとに、路上センサー整備済み路線

との相関を反映させながらリアルタイムに補完を行い、カバーエリアを拡大できる意味は大きいと考えている。

計算コストの点において、2 章に述べた特徴空間の生成は、固有値演算などから構成される PCAMD に数分程度の時間を要したが、これはオフラインで実施すればよい処理である。一方、リアルタイムに実施する必要のあるステップ 2, 3 は、行列積および逆行列演算のみから構成されており、高速な処理が可能である。実際、今回の評価では Pentium III-800 MHz の計算機上で、150 リンクの 1 週間分の補完処理、すなわち延べ 15 万リンクの演算を同時に行ったが、それに要した時間は 100 秒足らずである。

全国規模のサービスに際しては、数百リンク単位に固有の基底を用意する必要があるため、メモリに展開すべきデータ量は数十 MB 程度に増える。一方、演算量は延べリンク数に線形であり、処理周期ごとの演算対象リンク数は現状で数十万程度である。それゆえ、リアルタイム性を要する処理自体は実用上問題ない時間範囲で実現できる見通しであり、本技術の実サービスへの適用も十分に可能であると考えている。

5. 結 言

本研究では、プローブカーデータのリアルタイム補完において現況データを反映し、かつ、精度良く補完を行うために、特徴空間射影を用いた補完手法を開発した。今後はプローブカーシステムと本手法を組み合わせ、欠損のない網羅的な交通情報をユーザに提供する実用化フェーズの検討を進めていきたいと考えている。また、その情報を予測や経路誘導などの各種アプリケーションに応用することで、より魅力的な交通情報サービスを実現できるものと確信している。

そのような実用化の取り組みと並行して、特徴空間データ補完のアルゴリズムそのものについても、重み付けの定式化と動的制御などにより、さらなる推定精度の向上を図っていく考えである。また、いかなる条件で推定精度の良否が決まるのか、適用できる限界の欠損率はどのくらいか、理論と実験の両面から見極めていきたい。

参 考 文 献

- 1) Innamaa, S.: SHORT-TERM PREDICTION OF HIGHWAY TRAVEL TIME USING MLP-NEURAL NETWORKS, *Proc. 8th World Congress on ITS Sydney*, CD-ROM (2001).
- 2) 舟橋賢二ほか: 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, CD-ROM (2003).

- 3) Stutz, C., et al.: Classification and Prediction of Road Traffic Using Application-Specific Fuzzy Clustering, *IEEE Trans. Fuzzy Systems*, Vol.10, No.3, pp.297-308 (2002).
- 4) 熊谷正俊ほか：全国規模の交通情報サービスを目的とする所要時間長期予測技術の開発，情報処理学会論文誌，Vol.45, No.12, pp.2696-2705 (2004).
- 5) 柘植正邦ほか：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 (2004).
- 6) Miwa, T., et al.: En-Route Updating Methodology of Travel Time Prediction Using Accumulated Probe-Car Data, *Proc. 11th World Congress on ITS Nagoya*, CD-ROM (2004).
- 7) Fushiki, T., et al.: Study on Density of Probe cars Sufficient for Both Level of Area Coverage and Traffic Information Update Cycle, *Proc. 11th World Congress on ITS Nagoya*, CD-ROM (2004).
- 8) Ruhe, A.: Numerical computation of principal components when several observations are missing, Tech Rep. UMINF-48, Dept. Information Processing, Umea Univ. (1974).
- 9) 柴山 直ほか：欠損値がある場合の線形等化法，教育心理学研究，Vol.35, No.1, pp.86-89 (1987).
- 10) 高根芳雄：制約付き主成分分析法，朝倉書店 (1995).

(平成 17 年 11 月 28 日受付)

(平成 18 年 5 月 9 日採録)



熊谷 正俊 (正会員)

1975 年 12 月 2 日生まれ。2002 年 3 月東北大学大学院情報科学研究科システム情報科学専攻博士後期課程修了。博士 (情報科学)。同大学における脚車輪型移動ロボットの予測歩容制御の研究を経て，同年 4 月 (株) 日立製作所入社，日立研究所勤務。交通情報サービスの研究開発に従事。日本機械学会，日本ロボット学会各会員。平成 16, 17 年情報処理学会 ITS 研究会優秀研究報告賞，情報処理学会 DICOMO2005 優秀論文賞受賞。



伏木 匠 (正会員)

1973 年 10 月 21 日生まれ。1998 年 3 月東京大学大学院工学系研究科産業機械工学専攻修士課程修了。同年 4 月 (株) 日立製作所入社，日立研究所勤務。交通管制システムの開発を経て，現在，交通情報サービスの研究開発に従事。自動車技術会会員。平成 14 年情報処理学会 ITS 研究会優秀研究報告賞受賞。平成 16 年度電機工業技術功績者奨励賞受賞。



横田 孝義 (正会員)

1956 年 10 月 28 日生まれ。1984 年 3 月東京工業大学総合理工学研究科精密機械システム専攻博士後期課程修了，工学博士。同年 4 月 (株) 日立製作所入社，日立研究所勤務。1988 年より 1 年間米国カーネギーメロン大学客員研究員を経て，以降，交通管制システム，交通情報サービスの研究開発に従事。1993 年 (社) 交通工学研究会徳岡記念賞受賞。



君田 和也

1979 年 1 月 24 日生まれ。2003 年 3 月京都大学大学院情報学研究科知能情報学専攻修士課程修了。同大学におけるインタラクティブバーチャルスタジオの研究を経て，同年 4 月 (株) 日立製作所入社，日立研究所勤務。現在，交通情報サービスの研究開発に従事。