

プローブカーデータのリアルタイム補完技術

熊谷 正俊 伏木 匠 横田 孝義 君田 和也

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: {mkumagai, tfushiki, tyokota, kkimita }@gm.hrl.hitachi.co.jp

要約 プローブカーデータの欠損をリアルタイムに補完し、網羅的な交通情報を提供する技術について述べる。この補完技術は大きく分けて、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、欠損を含む現況データの特徴空間への射影、そして、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の間で相関のある成分ごとに欠損情報の推定が為されるため、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高精度に補完することが可能である。

キーワード プローブカー, 欠損値, 補完, 特徴空間射影

Realtime Interpolation Method for Probe Car System

Masatoshi KUMAGAI Takumi FUSHIKI Takayoshi YOKOTA and Kazuya KIMITA

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

E-mail: {mkumagai, tfushiki, tyokota, kkimita }@gm.hrl.hitachi.co.jp

Abstract This paper discusses about realtime interpolation method for probe car data, which provides traffic information with no missing data. The method consists of three major processes: determination of a feature space from past probe data history; feature space projection of current probe data; and 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, i.e., correlated component of a lot of links, the method can exactly estimate traffic condition caused by various factors.

Keyword Probe Car, Missing Data, Interpolation, Feature Space Projection

1. 緒言

渋滞を避けて快適に走りたい、少しでも早く目的地に行きたい、交通情報サービスの役割はドライバーのそうした希望に応じて、渋滞箇所を適確に知らせ、最適な経路へと導き、また、所用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遂げつつあり、その方向性は大きく分けて2つあると言える。ひとつは、予測情報の提供という時間方向の拡張である。予測情報の利用により、長距離の移動や数日先のドライブ、あるいは長期的な配送計画においても、適切な経路選択や所用時間の見積もりが可能になる。そしてもうひとつは、プローブカーによって交通情報の提供

エリアを拡大しようという、空間方向の拡張である。既存の路上センサーではカバーされていない道路についても交通情報を提供することで、より適切な経路選択が可能になる。

本報告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輛自体がセンサーであるため、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、原理的には全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、センサーである車輛の走行位置およびタイミングが確率的なものであることから、その情報品質は路上センサー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、プローブデータが取得できる時間密度は、道路リンク当たりの平均で1時間に1回程度である^[1]。これを現行の路上センサーと同等の5分周期、ある

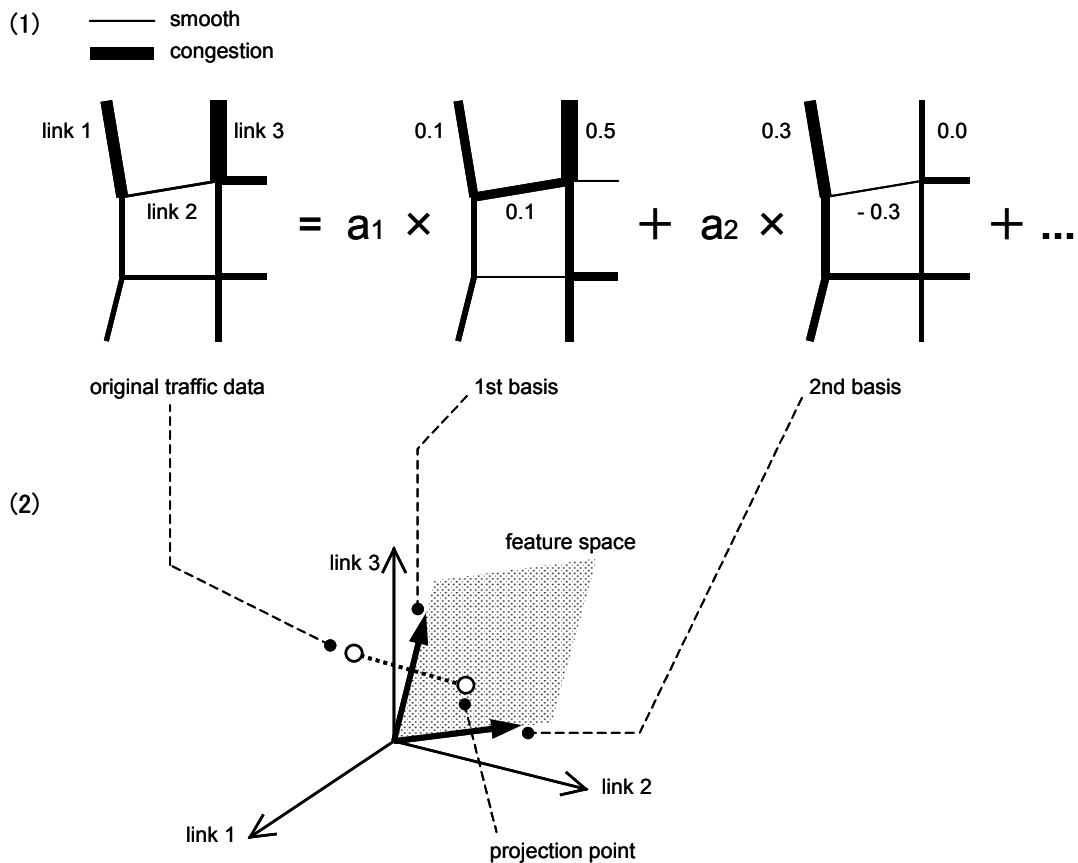


Fig.1 Traffic condition of correlated links represented by feature bases.

いは倍の10分周期のデータとして利用するにしても、時系列上での欠損率は8割から9割以上に達する。このように疎らなプローブデータを路上センサーと相補的な情報源として活用するには、何らかのデータ補間手段が必要となる。

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データから同時刻平均値を算出しておき、現況のプローブ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という手法である。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、曜日や季節による変化に対して十分に追従できないという問題がある(曜日変化等を反映するに過去データを詳細に分類する必要があるが、分類単位ごとのサンプルが減少することで、統計的な信頼性はむしろ低下する)。

この問題は、所要時間の見積も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、サービスの魅力を低減させる。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言える。

本報告では以上の観点から、新たに開発したプローブデータのリアルタイム補完手法について解説する。この手法は、過去のプローブデータから特徴空間を生成し、現況のプローブ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、本報告の構成である。2章では新たに開発したリアルタイム補完手法のアルゴリズムについて述べる。3章はその効果を評価した結果である。4章は結言であり、今後の課題、展望などについて述べる。

2. 特徴空間射影を用いたリアルタイム補完

2.1. リアルタイム補完の基本アルゴリズム

本手法の基本的な考え方を模式的に表すとFig. 1(1)のようになる。Fig. 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も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別に、“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の交通状況がどのような傾向にあるかを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情報量の次数によって決まる。

Fig. 1(1)の考え方を特徴空間という観点でとらえると、Fig. 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、その係数が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、その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Fig. 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、通常のペアワイズ、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。

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らによってそもそも提案された数値的な手法^[2]と、後年、柴山らによる欠損値を加味した線形等化法^[3]を高次元に拡張した解析的な手法^[4]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。 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 \mathbf{X} で表す。

\mathbf{X} の i 行目の成分を対角要素とするデータ行列 \mathbf{D}_{xi} 、重み行列 \mathbf{V} 、 \mathbf{V}_0 から成る

$$\mathbf{Y}_i = \mathbf{D}_{xi}\mathbf{V} + \mathbf{V}_0 \quad (1)$$

に対して、PCAMDによる特徴空間基底の決定は、制約条件

$$\mathbf{V}^T \mathbf{S} \mathbf{V} = \mathbf{I} \quad (2)$$

の下で、SSノルム

$$J = \sum_{i=1}^N \text{SS}(\mathbf{y}_i - \mathbf{e}_M \mathbf{u}_i^T)_{\mathbf{D}_{xi}, \mathbf{I}} \quad (3)$$

を最小化するという問題である。ここで、 \mathbf{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、 \mathbf{D}_{wi} は \mathbf{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、

$$\mathbf{Q}_{M/D_{wi}} = \mathbf{I} - \mathbf{e}_M \mathbf{e}_M^T \mathbf{D}_{wi} / \text{tr}(\mathbf{D}_{wi}) \quad (4)$$

$$\mathbf{C}_i = \mathbf{Q}_{M/D_{wi}}^T \mathbf{D}_{wi} \mathbf{Q}_{M/D_{wi}} \quad (5)$$

$$\mathbf{A}_1 = \sum_i \mathbf{D}_{xi} \mathbf{C}_i \mathbf{D}_{xi} \quad (6)$$

$$\mathbf{A}_2 = \sum_i \mathbf{D}_{xi} \mathbf{C}_i \quad (7)$$

$$\mathbf{A}_3 = \sum_i \mathbf{C}_i \quad (8)$$

$$\mathbf{A} = \mathbf{A}_1 - \mathbf{A}_2 \mathbf{A}_3 + \mathbf{A}_2^T \quad (9)$$

に対して、固有値問題

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

を解くことによって得られる。ただし、

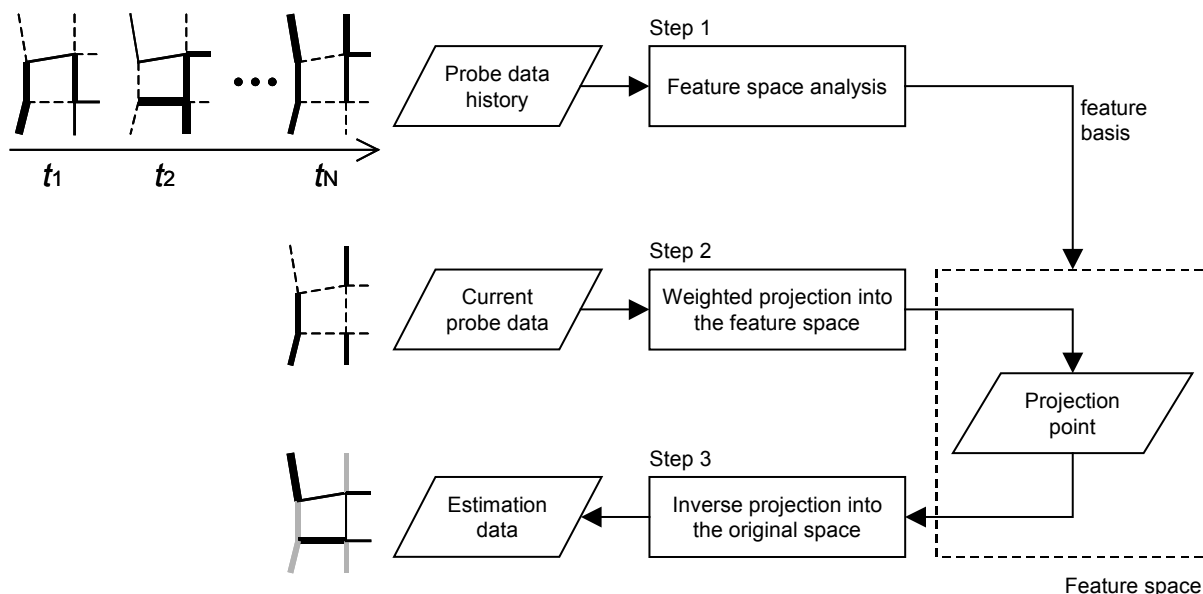


Fig.2 Interpolation process of probe car data.

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

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

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

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基底行列 P に対して、現況のプローブデータ 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。また、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、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 の補完が為される。

2.2. 重み付けによる現況データ利用の拡大

前節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、過去における突発的な事故や工事による渋滞などの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。

以上の問題点を解消して、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、前節ステップ2において、特徴空間に射影する現況データの時間幅を長くとり、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、わずかな時間で交通状況が大きく変化することもあることから、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するため、式(14)の重み付け射影における重み行列 W を、現況データのタイムスタンプに応じた可変値とすることを考えた。これは、重み行列 W の要素 w_j を、たとえば次式のような現在時刻 t とデータ収集時刻 t_c の関数として、式(14)の重み付け射影を解くという手法である。

$$w_j = \begin{cases} z_1 & (0 \leq t - t_c < T_1) \\ z_2 & (T_1 \leq t - t_c < T_2) \\ z_3 & (T_2 \leq t - t_c < T_3) \\ 0 & (T_3 \leq t - t_c, \text{ missing}) \end{cases} \quad (16)$$

$$0 < z_3 < z_2 < z_1, \quad 0 < T_1 < T_2 < T_3$$

このようにタイムスタンプに応じた可変値を用いて重み付けをすれば、古いデータも活用して特徴空間座標を安定に決定するための観測データを確保し、かつ、新しいデータをより重視して補完を行うことが可能になる。

3. 補完精度の評価

本報告では実際のプローブデータの代わりに、路上センサーで収集されたリンク旅行時間を人為的に欠損させ、そのデータを用いて提案手法の検証を行う。現時点では路上センサーデータのほうが潤沢であり、広域的な評価に適していること、また、評価条件に応じた欠損パターンを用意することが可能なためである。プローブ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。

評価方法は下記のとおりである。

- (1) ある期間のリンク旅行時間をランダムに欠損させ、それに基づいて2.1節ステップ1で述べた特徴空間の生成を行う。(以下では特徴空間の生成に用いるデータを“学習データ”と呼ぶものとする。)
- (2) 続いて、学習データより後の期間のリンク旅行時間を、学習データとは異なるパターンでランダムに欠損させ、2.1節ステップ2に従って(1)の特徴空間に射影する。(特徴空間に射影するデータは“評価データ”と呼ぶものとする。)
- (3) 2.1節ステップ3に従って、(2)の特徴空間座標を元の交通情報データ空間に逆射影し、評価



Fig.3 Evaluation area.

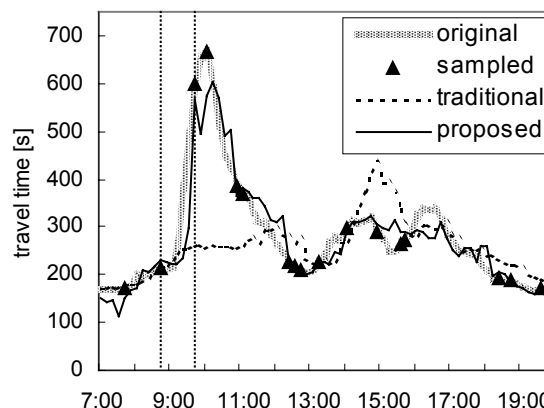


Fig.4 Example of interpolation.

データの欠損箇所について、逆射影された値を補完値とする。

以上により得られた補完値を、評価データ欠損箇所の真値と比較し、時刻ごとに全リンクの補完誤差の平方二乗平均(RMS)をとった値が評価基準である。

次に評価条件を示す。評価にはFig. 3の枠内における主要150リンクのリンク旅行時間を用いた。上記学習データの期間は2001年4月1日～7日、評価データの期間は同8日～14日である。なお、路上センサー本来のデータ周期は5分だが、データが大量であるため便宜的に等間隔に半分の間引き、10分周期のデータとして扱った。

Fig.4 は実際に補完処理をして得られた推定値を、1本のリンクについて真値および従来手法と比較したグラフである。このグラフにおいては、午前中に通常とは異なる傾向の渋滞が生じ、従来手法による同

時刻平均値とは大きく乖離している。一方、提案手法を用いれば、現況の観測値を反映した推定が可能であり、特に9～10時の観測値が全くない時間帯においても、渋滞の立ち上がりをよくとらえていることがわかる。これが周辺リンクとの相関に基づいて補完を行う効果である。続いては、定量的な評価結果を示す。

特徴空間射影を用いる補完の効果

Fig. 5 (a)は、従来の同時刻平均を用いた手法と、本報告で提案する補完手法とについて、時刻ごと補完誤差RMS値の分布を比較したヒストグラムである。グラフでは横軸が補完誤差RMS値の階級を表し、縦軸はその度数である。従来手法では1.0程度まで広く分布していたRMS値が、0.2～0.5程度に縮減されていることがわかる。

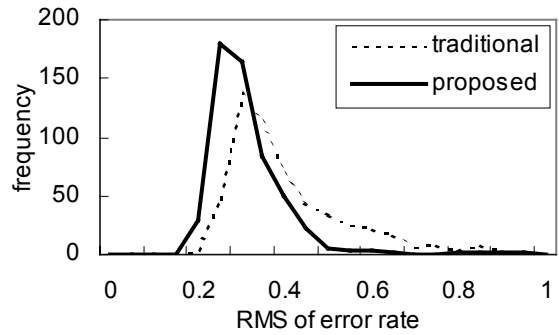
現況データの時間幅拡大の効果

Fig. 5 (a)では提案手法における特徴空間の次元を5次元としていたが、Fig. 5 (b)は、それを特徴空間次元10次元の場合と比較したヒストグラムである。図中の凡例において、Dは特徴空間の次元を表す。特徴空間次元を増やしたことで、誤差が若干増加していることがわかる。

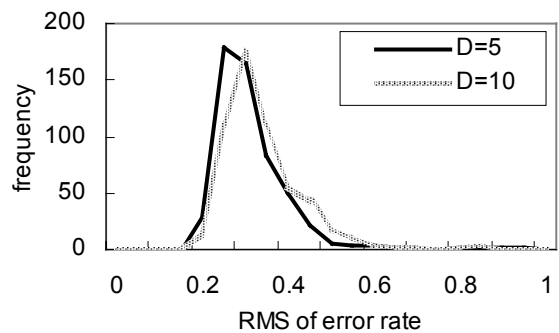
これに対して、Fig. 5 (c)は、2.2節に述べた可変値の重み付けにより、現況データとして利用する時間幅を拡大した結果のヒストグラムである。現在時刻から20分前のデータまでを現況データとして利用し、タイムスタンプによる重み付けは10分刻みに1.0, 0.8, 0.6とした。図中の凡例において、Dは特徴空間の次元を表し、Tは現況データの時間幅を表す。この結果、最も補完精度が高いのは特徴空間次元5次元、現況データ時間幅20分とした場合だが、特徴空間次元を10次元としても、5次元の場合と遜色ない良好な補完データが得られたことがわかる。

Table 1は、以上の補完誤差の比較において、各評価対象におけるRMS値の平均値を列記したものである。これらの結果を総合すると、特徴空間を用いた補完を行い、また、可変重み付け射影を行って現況データの時間幅を拡大することで、従来手法に比して補完誤差を2/3に低減できたということが結論として言える。

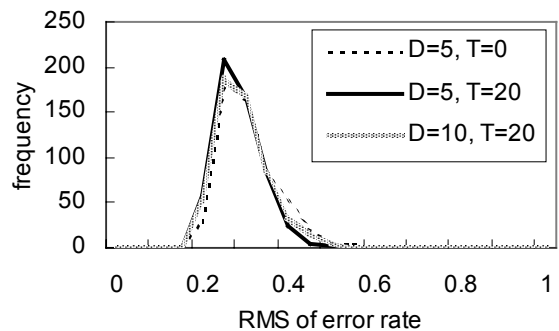
計算コストの点においては、オンラインで実施する必要のある2.1節ステップ2, 3の処理は、行列積および逆行列演算のみから構成されており、高速な処理が可能である。また、一度に数百リンク単位の補完処理が可能であるため、全国規模のリアルタイムサービスにも直ちに適用が可能である。実際、今回の評価ではPentiumIII-800MHzの計算機上で、150リンクの1週間分の補完処理、すなわち延べ15



(a)



(b)



(c)

Fig.5 Evaluation result.

Table 1 Comparison of interpolation error.

	traditional	proposed (D=5)		proposed (D=10)	
		T=0	T=20	T=0	T=20
RMS	0.383	0.287	0.260	0.312	0.268

万リンク分の処理を同時に行ったが、それに要した時間は100秒足らずである。

4. 結言

本研究では、プローブ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、精度よ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ーに提供する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。

参 考 文 献

- [1] T. Fushiki, et al., “Study on Density of Probe Cars Sufficient for Both Level of Area Coverage and Traffic Information Update Cycle,” Proc. of 11th World Congress on ITS Nagoya, CD-ROM, Japan, Oct. 2004.
- [2] A. Ruhe, “Numerical computation of principal components when several observations are missing,” Tech Rep. UMINF-48, Dept. Information Processing, Umea Univ., 1974.
- [3] 柴山, “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.
- [4] 高根, “制約付き主成分分析法,” 朝倉書店, 1995.