

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

熊谷正俊[†] 伏木 匠[†] 横田孝義[†]
佐野 豊^{††} 鈴木研二^{††}

本論文では全国規模の交通情報サービスを目的とした所要時間長期予測手法を提案する。この手法は曜日/祝祭日/季節/五十日等の多様な因子を加味した予測が可能であり、全国域の大規模なデータをコンパクトな予測モデルで取り扱うことができる。その基本となる考え方は、たとえば朝夕の渋滞に相当するような複数の基底データが日付に応じた強度で合成され、所要時間データを構成しているというものである。この考え方に基づけば、所要時間データは日付によって変化しない基底データと、日付によって変化する合成強度とに分解され、曜日等の因子と関連づけて予測すべき情報は後者のみである。基底データと合成強度への分解には、主成分分析を用いた特徴空間射影が利用可能であり、合成強度に関する予測は特徴空間上で行われる。所要時間データ空間に比べて低次元化されていることから、特徴空間上では予測モデルを肥大化させることなく、高精度な予測処理を行うことが可能である。特徴空間からの逆射影は分解された基底データの合成による所要時間データの復元に相当し、特徴空間上で予測された合成強度の逆射影により所要時間データの予測値が得られる。実際の所要時間データを用いた評価により、この手法が予測精度の効率的な改善に有効であることが確認された。

Development of Long-term Travel Time Forecast Method for Nationwide Traffic Information Services

MASATOSHI KUMAGAI,[†] TAKUMI FUSHIKI,[†] TAKAYOSHI YOKOTA,[†]
YUTAKA SANO^{††} and KENJI SUZUKI^{††}

We propose a travel time forecast method for nation wide traffic information services. This method can deal with various “day factors” which traffic condition depends on, e.g. days, seasons, and vacations, without complicated local parameter setting. It also has the advantage in both the size of the database and the amount of the calculation for forecasting. The basic idea of the method is that travel time data consists of the weighted sum of several “feature bases” which represent characteristic traffic patterns such as congestion in morning or evening. The feature bases are the information which is independent of the day factors; in contrast, the information to be forecasted according to the day factors is weighting coefficients of the feature bases. The feature bases are given by the feature space projection using principal component analysis, and the forecasting process for the weighting coefficients is performed in the feature space. Since the dimension of the feature space is smaller than that of the travel time data space, the method can achieve an accurate forecasting process with a small database and a little amount of calculation. The information of the feature bases are restored to the forecast data of travel time by the inverse projection from the feature space. Evaluation results show that the method efficiently improves the forecast accuracy for wide-area applications.

1. 緒 言

これから走る道は空いているのか混んでいるのか、渋滞はいつどこで終わるのか、どの道を通れば目的地まですなりと行けるのか、ドライバなら誰もが気

にかける事柄である。こうした疑問に応える情報サービスは、これまで主に行政の事業として整備が進められてきた。全国の主要道路で収集された交通情報はVICS 対応カーナビ等にリアルタイムに配信され、高速道路についてはインターネット上で予測情報の提供も行われている。

一方で交通情報の利用には、交通情報に基づく経路案内、物流業務への活用など、さらに多様なサービスへのニーズが存在し、それらは人や物の移動に要す

[†] 株式会社日立製作所日立研究所

Hitachi Research Laboratory, Hitachi Ltd.

^{††} 株式会社日立製作所情報制御システム事業部

Information & Control Systems Division, Hitachi Ltd.

るコストを抑え、環境負荷を緩和し、より快適な道路交通を実現するものとして期待されている。そうしたニーズを背景に、2002年に施行された交通情報に関する規制緩和を契機として民間企業による交通情報ビジネスが開始され、2004年3月の時点で、全国域の予測情報を提供する携帯電話コンテンツサービスや、交通状況を反映した経路誘導を行うカーナビ等がすでに実用化されている。

これらのサービスは、道路管理者が収集し、VICSセンターが集約/配信するVICS情報を独自に加工し、携帯電話等の通信回線を通して、より付加価値の高い情報をオンデマンドにユーザに提供している。また、交通情報の収集を独自に行う取組みもすでに開始されている。このような新しい交通情報サービスは、利便性の高い通信システムの普及によってもたらされたものでもあり、他のテレマティクスサービスとともに、今後もさらに発達を続けていくだろう。

筆者らの研究グループも、ドライバや物流事業者のニーズに応えるサービスの提供を目的として、交通情報の予測、プローブカーによる交通情報収集¹⁾、安全診断²⁾等の技術開発を行っている。その中で本論文において述べるのは所要時間の長期予測技術である。

たとえばドライブの計画を立てるときに、週末や長期休暇の道路状況を前もって把握しておくことができれば、時間の見積りやルートを選択が容易になり、より快適にドライブを楽しむことができる。物流事業の分野においては、現状では配送計画における時間的マージンを大きく見積もることで遅延リスクに対応しているが、予測情報を利用した遅延リスクの定量化により、運行コストの削減が可能になると考えられるケースもある。社会的に見れば、道路交通の中で少なからぬウェイトを占める物流車輛の運行を効率化することで、渋滞の緩和、環境負荷の低減などの効果が期待される。筆者らはこのような観点から、早期の実用化を強く意識して所要時間長期予測技術の開発を進めてきた。この技術はすでに所要時間予測エンジンとして製品化され、テレマティクスコンテンツプロバイダによる交通情報サービスのシステムに組み込まれ、一般ユーザに利用されている。

実用に供する予測技術として、開発にあたり留意したのは次の各要件である。

- (1) 交通状況に影響を及ぼす多様な因子を予測条件として加味する。
- (2) 全国域の情報提供に対応する。すなわち、地域・路線ごとのチューニングを要するパラメータはなるべく排除する。

- (3) 予測処理に用いるデータベースのサイズを抑える。

(1)は予測情報の提供を行ううえで第1に目指した点である。たとえば曜日による違い、季節による違い、あるいは、地域によっては商用車の交通量に少なからぬ影響を及ぼすといわれる五十日など、それらを予測条件として設定できるようにする。一方、(2)は(1)とのトレードオフになるが、全国域の交通情報を扱う立場で、多様な地域の情報を確実に提供するには、地域ごとに複雑なチューニングを施すのではなく、全国一律のパラメータ設定によって十分な予測精度を確保することが必要である。むしろ、地域カスタマイズをすることはサービスのクオリティ向上に効果的なので、そのための仕組みもあわせて用意しておく必要がある。

(3)は個々の道路リンク単位ではさしたる問題ではなくても、全国規模の情報を扱ううえでは、VICSリンクとして定義されたものだけで全国に24万本近くあるというスケールで作用する問題であり、予測エンジンの性能、動作環境、あるいはデータベース更新の時間など、システム運用の観点から無視することのできない要素である。

本論文では以上の論点を中心に、開発した所要時間長期予測技術の報告を行う。本論文の構成を下記に示す。2章では要件(1)~(3)を満たし、全国規模の予測情報サービスを実現するうえでの課題について明らかにする。3章はその課題を解決するために開発した手法の解説である。4章では実際の道路の所要時間データに基づいて、開発した手法を評価した結果を示す。5章は結言であり、関連技術やこれからの課題について整理する。

2. 所要時間長期予測の課題

所要時間を含め、渋滞度や交通量など交通情報の予測は、過去において交通情報にどのような変化パターンが現れたか、それが未来においてどのように再現されるか、基本的にはその統計的な関連付けによって行われる。これは、交通状況が1日を最小単位として周期的変化をすることに立脚したアプローチである。

短期予測：数時間程度のスパンで交通情報を予測する場合には、現況と予測対象期間との間に一般に強い相関があるため、過去の交通情報の履歴から、現況の交通情報に基づいて予測情報を生成する。具体例としては、回帰分析³⁾、ニューラルネット⁴⁾、パターンマッチング^{5),6)}等の手法を用いて、現況を基に未来の交通情報を高精度に推定する技術が報告されている。

一方、交通情報を長期的に予測する場合には、現況

と予測対象期間との相関が低く、短期予測のように現況の交通情報に基づいて予測情報を生成することはできない。たとえば、ある日の道路の混み具合が平均的な状態を上回っているからといって、その翌日、さらには1週間後も同じように混雑するのか、容易には判断し兼ねる。しかしながら、「金曜日だから混んでいる」「観光シーズンだから混んでいる」といった具合に、その混雑が何らかの因子によって理由付けされるものであるならば、1週間後でも1カ月後でも、その因子に基づく予測が可能である。これが長期予測の基本的な考え方であり、曜日/祝祭日/五十日/連休/月/季節/イベントなどの因子（これらを“日種因子”と呼ぶ）と交通情報との統計的な関連付けによって、過去の交通情報の履歴から、未来の交通情報を予測する。

日種因子に基づく長期予測技術として一般的であり、実用化の実績もある手法は、日種因子をラベルとして過去の交通情報の時系列データを1日分ずつ分類し、分類された集合ごとに同時刻平均値等の統計的代表値を予測情報として提供する、日種分類手法である⁷⁾。1章で述べた要件(1)に関して、ある具体的な日付には複数の日種因子が対応していることから、日種分類を行ううえでも“10月:月曜日:五十日”のように日種因子を組み合わせれば、予測精度の向上に有効であると考えられる。

一方で、日種因子を組み合わせることに、分類された集合ごとのサンプル数が減少し、予測の信頼性がむしろ低下するという問題がある。たとえば、過去1年間に収集されたデータから予測を行うものとして“10月:月曜日”のように月と曜日の因子を組み合わせた場合、そのサンプル数は4か5である。因子を追加すればサンプル数はさらに少なくなり、サンプルがまったく得られないこともある。クラスタリングは日種分類を最適化し、多様な日種因子の組合せに基づいて予測精度を向上するのに有効な手法だが⁸⁾、一般に複雑なパラメータ調整等を要するため、要件(2)の点において、多様なデータを取り扱う全国域の交通情報サービスに適用することは難しい。

日種分類手法が細分化されたサンプル集合から予測データを計算するのに対して、回帰分析を用いれば、交通情報データを日種因子の関数として表現し、その関数を用いて予測を行うことができる。回帰モデルのパラメータは、個々の日種因子と交通情報データの全サンプル集合との相関に基づいて決定されるので、多様な日種因子を加味した信頼性の高い予測が可能である。非数値情報を説明変数とする数量化I類は、非線形な成分は扱えないものの、回帰モデルの同定が容易

であり、交通情報の分析/予測にも広く用いられている。多重共線性等に注意すれば、説明変数の選択に関する自由度も高い。

しかしながら、全国規模の交通情報を取り扱う用途においては、回帰分析の利用にも制約がある。それは要件(3)で述べた予測データベースのサイズに関する問題である。たとえば、筆者らは全国共通の日種因子として約50種類の因子を予測に用いており、これに加えて約30種類の因子の利用を検討中である。全国で約24万本定義されているVICSリンクのうち、一部については情報の提供はなされていないが、その整備は継続して進行中である。そこで、約半数の12万本のリンクについて予測を行うものと仮定し、予測データの時間分解能をVICSと同等の5分とすると、そのデータベースのサイズは11Gバイトにも達する。

この予測データベースをメモリ上に展開して扱うには相当にハイエンドな計算機が必要である。かといってHDD上に置いたのではアクセスに時間がかかり過ぎ、動的経路探索のような用途ではとうてい使いものにならない。カーナビでの利用ともなると、交通情報だけで記憶領域を消費するわけにはいかないので、実装はなおのこと難しい。地図データに比べれば交通情報の空間分解能ははるかに低いが、時間軸方向の成分を持つために、取り扱うべきデータ量は決して少なくないのである。

データ量の問題は計算機性能の向上によって解消される性質の事柄ではある。しかし、一方では、プローブカーを用いた交通情報収集の取組み等がITS業界において着々と進められており、将来的にはさらに広範囲にわたる詳細なデータをハンドリングするニーズが生じるのは必然である。それゆえに、交通情報サービスのこれからの方向性として、高精度かつコンパクトな予測技術は必須開発項目の1つであると考えている。

3章で解説するのは、以上に述べた日種因子の多様性、信頼性、データ量などに関する課題を解決する所要時間予測手法であり、ベースとなる予測技術は回帰分析である。回帰分析をベースとしたのは、信頼性が高く、回帰モデルの工夫で非線形な成分にも対応できること、過去に同一の事例がなくても日種因子の組合せで類推ができることなど、拡張性と汎用性が高いと考えたためである。そのうえでデータ量の問題もクリアするために、所要時間データを日種因子によって変化しない情報と、変化する情報とに分解し、後者のみを回帰分析の対象にするというアプローチをとった。

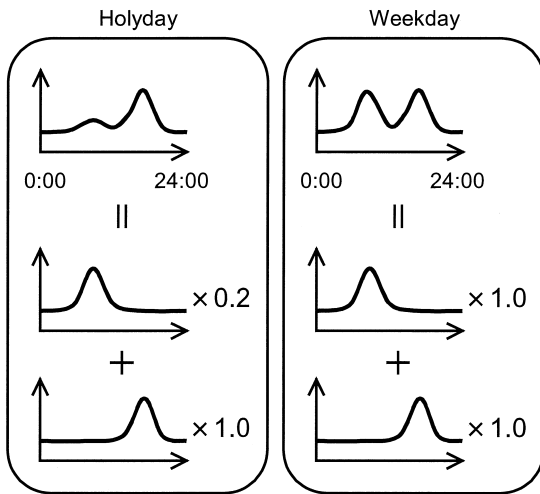


図 1 特徴基底による所要時間データの合成

Fig. 1 Travel time data represented by the feature bases.

3. 特徴空間予測手法

本研究で基本とする考え方は、所要時間データが複数の“特徴基底”の線形合成によって構成されているというものである。特徴基底とは朝ピークや夕ピークなど、予測対象リンクの所要時間データに特徴的なパターンを表し、1日分の長さを持つ時系列データである。たとえば図1のように、休日の所要時間データは朝ピークを表す特徴基底が強く、夕ピークを表す特徴基底が強く合成されたものであり、一方、平日のそれは、朝ピークと夕ピーク双方の特徴基底が強く合成されたものであるといえる。

ここで着目すべきは、「休日の朝は空いて夕方混雑する」「平日は朝と夕方が混雑する」という具合に、日種因子に関連した交通状況の特徴によって、特徴基底の合成強度が異なる点である。すなわち、合成強度は日種因子の関数である。一方の特徴基底は、日付によってその強度が変わるのであって、それ自体はリンク固有の不変な情報として定義可能と考えられる。特徴基底の情報が分離された分、合成強度の次数は所要時間データに対して低くなる。ゆえに、適切な特徴基底を定義することができれば、日種因子に基づく予測対象データの低次元化が可能になる。

この考え方に基づけば、次のプロセスに沿って図2のように所要時間データを予測できると考えられる。プロセス1 過去の所要時間データから特徴基底を抽出する。

プロセス2 過去の所要時間データにおける特徴基底の合成強度と、日種因子の履歴から、合成強度関

数を同定する。

プロセス3 予測対象日の日種因子から、プロセス2で同定した合成強度関数を用いて合成強度を算出し、それを係数としてプロセス1で抽出された特徴基底を合成する。

プロセス1, 2が予測のためのデータベースをオフラインで作成する処理であり、プロセス3が予測情報の提供に際して、予測データを生成する処理である。以下、プロセス1~3を解説する。

プロセス1：特徴基底の抽出

過去の所要時間データから特徴基底を抽出するうえで、その線形合成によって元の所要時間データを適切に表現するには、次の条件がある。

- 所要時間データが持つ互いに無相関な成分は、個別の特徴基底で表現する。
- 一方、所要時間データが持つ互いに相関のある成分は、単一の特徴基底で表現する。
- 特徴基底が所要時間データに対して持つ情報量が明確である。

これらはいずれも所要時間データの情報を集約し、必要に応じて情報量の少ない特徴基底を省略して、合成強度の次数を下げるためである。このような条件を満たす手法として、主成分分析があげられる。

主成分分析は多変量データを複数の正規直交ベクトル（結合係数ベクトル）を主軸とする部分空間に射影する手法である。この部分空間は次のような性質を持つ。

- 各主軸上に射影されたデータ（主成分得点）は互いに無相関である。
- 第1主軸上のデータ分散は、元のデータ空間内で最大である。また、第2主軸上の分散は、第1主軸を除いて最大である。以下同様に、第 m 主軸上の分散は、第1~第 $m-1$ 主軸を除いて最大である。

つまり、主要な結合係数ベクトルによって張られる部分空間は、情報損失量を最小にしつつ元のデータを近似表現する特徴空間である。特徴空間上への射影点は、主成分得点を係数とする結合係数ベクトルの線形合成で与えられることから、結合係数ベクトルが特徴基底として適当であること、また、主成分得点が特徴基底の合成強度に相当することが分かる。主成分分析を用いた特徴基底の算出手順を以下に示す。

1日あたり M 回、 N 日間にわたって同時刻に計測された所要時間データを $N \times M$ 行列 X で表す。所要時間データの同時刻平均値を表す M 次元行ベクトル \bar{x} 、全要素を1とする N 次元列ベクトル e を用い

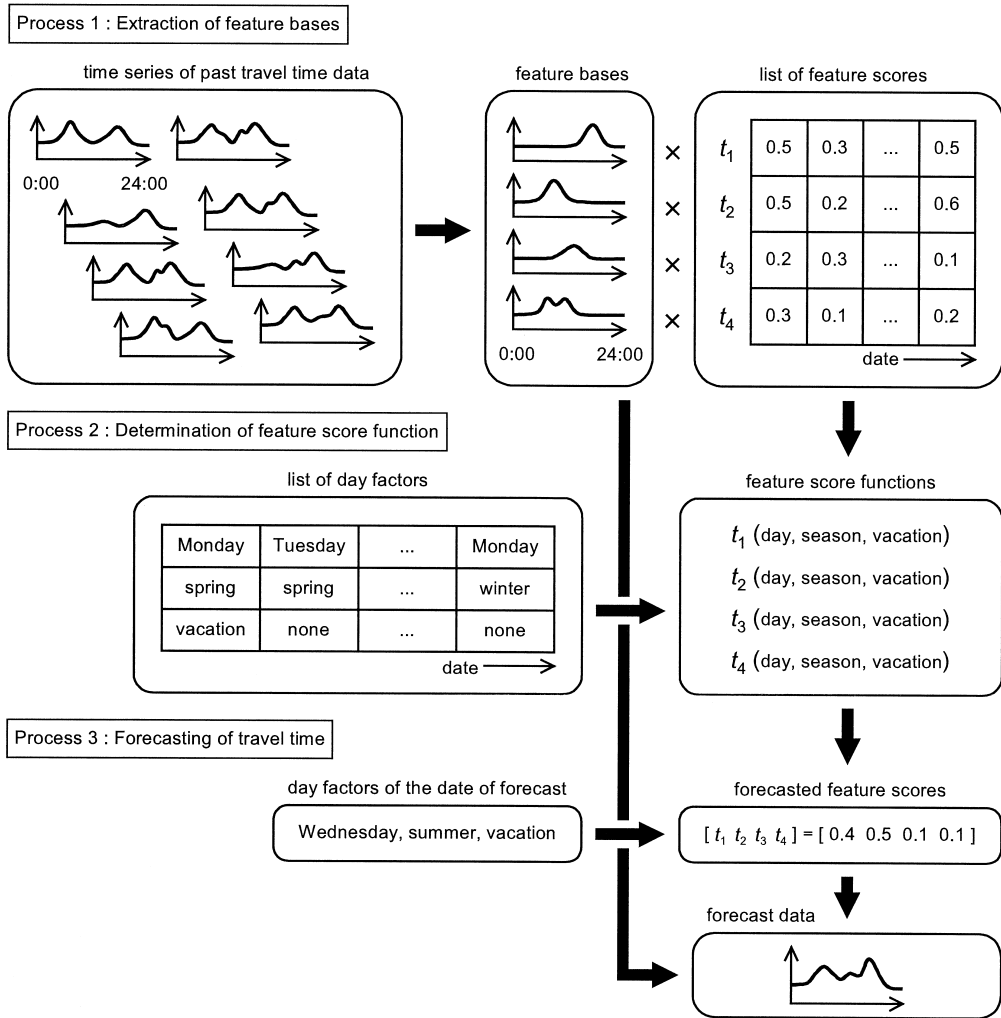


図 2 特徴空間予測手法の概念
Fig. 2 Basic idea of the proposed forecast method.

て、次式により X を列中心化する。

$$X^* = X - e\bar{x} \tag{1}$$

列中心化された所要時間データ X^* の列ごとの共分散

$$v_{ij} = \frac{1}{N-1} \sum_{n=1}^N X^*_{ni} X^*_{nj} \tag{2}$$

を計算し、 v_{ij} を要素とする共分散行列を V とする。 V の固有値問題

$$(V - \lambda_j I) w_j = 0 \tag{3}$$

を解いて得られる固有ベクトル w_j が結合係数ベクトル、すなわち特徴基底であり、対応する固有値 λ_j (主軸上の分散に等しい) が大きい順に p 個の固有ベクトルを並べた

$$P = [w_1 \ w_2 \ \dots \ w_p] \quad (p < M) \tag{4}$$

が特徴基底行列である。特徴基底の数 p は累積寄与率等から決定する。特徴基底行列を用いた射影

$$T = X^* P \tag{5}$$

で得られる T が主成分得点行列であり、その要素 t_{ij} が日付 i の所要時間データに関する特徴基底 j の合成強度に相当する。

図 3、図 4 はそれぞれ実際の所要時間データから得られた特徴基底と合成強度の事例を、主軸 1~5 について示したグラフである。主成分分析は必ずしも特徴基底に明示的な意味を与えることを保証しないが、朝ピークや夕ピークなど所要時間データにおいて顕著な変化を示す成分は図 3 のように明確に現れることが多い。同様に図 4 の合成強度には、平日/休日に対応した 1 週間周期の変化と、大型連休に相当する期間の変化が如実に現れている。元の所要時間データは、図 4

の合成強度を係数として、図3の特徴基底を線形合成することで近似される。

プロセス2：合成強度関数の同定

ここでは関数同定に回帰分析の1つである数量化I類を用いる。これは、表1のように日付と対応した日種因子を下式(6)の説明変数 d_i 、プロセス1で得られた合成強度を被説明変数 $t(d)$ として、 d_i による $t(d)$ の近似が二乗誤差最小となるように係数 a_i を決定するという問題である。

$$t(d) = \sum_i a_i d_i + b \tag{6}$$

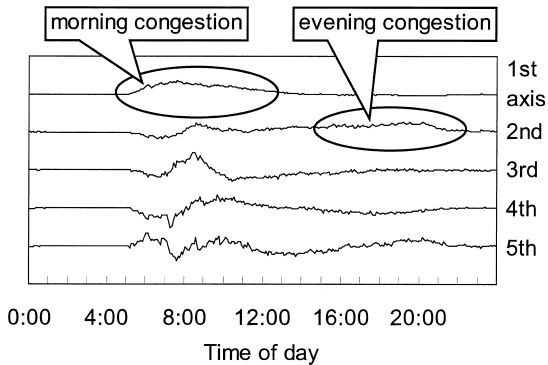


図3 特徴基底の事例
Fig. 3 Example of the feature bases.

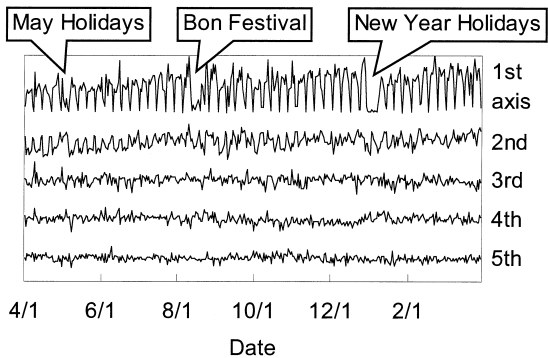


図4 合成強度の年間変化の事例
Fig. 4 Example of the component scores.

この問題は以下によって解かれる。

日付と日種因子の対応を日種因子行列 D として定義する。

$$D = \begin{bmatrix} 1 & 0 & 0 & 1 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 1 & 1 & 0 & \cdots & 1 & 1 \end{bmatrix} \tag{7}$$

D の定義において、要素 $D_{ik} = 1$ は日付 i が日種因子 k に該当することを表し、 $D_{ik} = 0$ は該当しないことを表す。プロセス1で得られた主成分得点行列 T の第 j 列 τ_j 、日種因子行列 D 、特徴基底 j の合成強度関数の係数ベクトル a_j について、誤差の二乗が最小となるように次の方程式を解く。

$$\tau_j = D a_j \tag{8}$$

そのような解は左側擬似逆行列を用いて

$$a_j = (D^T D)^{-1} D^T \tau_j \tag{9}$$

により計算される。これを全主軸について求めた

$$A = (D^T D)^{-1} D^T T \tag{10}$$

が合成強度関数の係数行列である。

プロセス3：予測データの生成

日種因子行列 D と同様の形式で予測対象日の日種因子を表す日種因子ベクトル d に対して、合成強度の予測値は、式(10)で求めた係数行列 A を用いて

$$\hat{t} = dA \tag{11}$$

により得られる。これを係数として特徴基底行列 P の各列ベクトルを合成した次式が、所要時間の予測データである。

$$\hat{x} = \hat{t}P^T + \bar{x} \tag{12}$$

先に述べたように、主成分分析で得られる結合係数ベクトルによって張られる部分空間は、元のデータに対する特徴空間である。たとえば図4に示した主成分得点のうち、第1、第2主軸をそれぞれ横軸、縦軸とする分布図をプロットすると図5のようになり、データの特性が顕著に表れていることが分かる。上記一連

表1 日種因子変数の事例

Table 1 Example of day factor variables.

day factors	day			season				vacation	
	Weekday	Saturday	Sunday	spring	summer	autumn	winter	vacation	none
variables	d1	d2	d3	d4	d5	d6	d7	d8	d9
2004/10/1	1	0	0	0	0	1	0	0	1
2004/10/2	0	1	0	0	0	1	0	0	1
2004/10/3	0	0	1	0	0	1	0	0	1
2004/10/4	1	0	0	0	0	1	0	0	1
2004/10/5	1	0	0	0	0	1	0	0	1

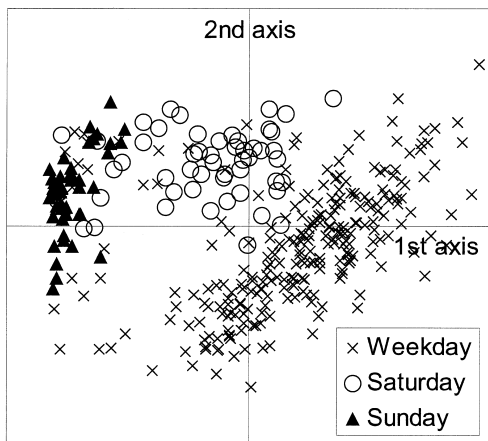


図 5 特徴空間における主成分得点の分布
Fig. 5 Scatter plot of the component scores.



図 6 評価対象区間
Fig. 6 The route for evaluation.

の予測プロセスにおいて、プロセス 1 が特徴空間への射影であるのに対して、プロセス 3 は特徴空間から所要時間データ空間への逆射影に相当する。このとき、日種因子によって変化しない成分は射影行列である特徴基底に保存され、逆射影の過程で再び取り出される。一方で、日種因子によって変化する成分に関わる予測処理はすべて低次元化された特徴空間上で実施される。このことから本手法を特徴空間予測手法と呼ぶことにする。

4. 評価

4.1 単一路線上での予測精度評価

特徴空間予測手法を用いて、多様な日種因子に基づく高精度な所要時間予測が可能であることを、実際の所要時間データを用いて評価する。評価対象とした道路区間は、東京都環状 8 号線の荻窪～等々力間 12.6 km の区間である (図 6)。この区間は北関東方面から東名高速や東京湾アクアラインに向かう商用車の利用が多く、予期せぬ渋滞が多いことで知られている。評価に用いたデータは道路交通情報システムで収集された所要時間データであり、2001 年度 1 年分のデータを基に、2002 年度のデータを予測した。

所要時間の予測事例を時系列データのグラフで図 7 に示す。この事例はごくふつうの平日 (2002 年 4 月 23 日) のものであり、図 7 では細線が実測データ、淡い太線が 2 章であげた日種分類手法による予測データ、濃い太線が特徴空間予測手法による予測データを表す。日種分類手法で用いたのは一般的な 3 種類の日種因子 (平日/土曜日/日曜日・祝祭日) である。特徴空間予測手法では平日/休日/祝祭日/五十日/連休/月など約 50 種類の日種因子を用い、特徴空間次元数を 30

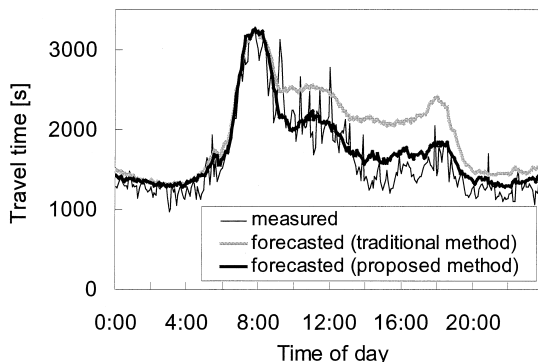


図 7 所要時間予測データの事例
Fig. 7 Example of forecasted travel time data.

次元とした。

図 8 はこの路線上での予測精度を定量的に評価した結果である。これは 2002 年 4 月～7 月の 4 か月間にわたって 5 分間隔に所要時間を予測し、その予測誤差の分布について、6 時～20 時台の各時間帯ごとに 80 パーセント誤差率を集計したグラフである。予測誤差率の定義は “ 実測値と平均値の差分の絶対値 / 実測値 ” である。80 パーセント誤差率とは、評価対象データの 80 パーセントが分布する誤差範囲の上限値であり、平均値よりも厳しい指標だが、情報の利用者にとっての当たる当たらないという感覚はこのあたりの水準にあるのではないかと思う。

両手法とも朝ピークの時間帯には実測データに対する誤差が小さいが、日種分類手法の場合、10 時以降は精度の悪化が著しく、80 パーセント誤差率は最大 45% に達する。これは商用車が多く走行する時間帯であり、年間を通じて変動の少ない通勤車輻に比

して、商用車の利用は季節や暦に応じて大きく変化するため、それに起因する交通状況を上記3日種因子のみでは十分に表現できていないのだと考えられる。一方、特徴空間予測手法の場合には、当該時間帯における誤差率が30%以下に低減されており、これが特徴空間予測手法を用いて多様な日種因子を利用することの効果である。

4.2 広域予測精度とシステム運用性の評価

ここでは広域的な予測精度と、システム運用性に関する項目を定量的に評価する。まず、特徴空間回数に関する評価を行う。表2は東京都内の3,500リンクについて所要時間データの主成分分析を行い、特徴空間回数と累積寄与率との関係をまとめた表である。この

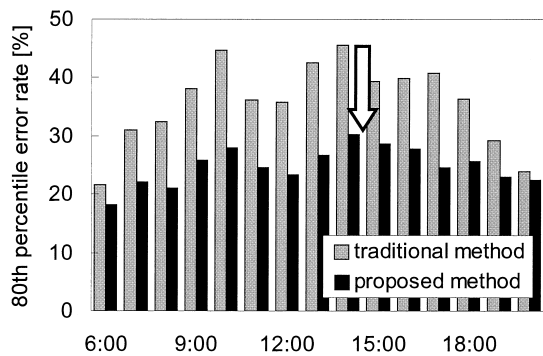


図8 1時間ごと予測誤差率(4カ月間の80パーセンタイル値)
Fig. 8 Forecasting error rate for every hour (80th percentile through four months).

表からは、特徴空間回数がおおむね20次元あれば元の所要時間データを十分に近似できることが分かる。たとえば、特徴空間回数を20次元としたときの80パーセンタイル値は0.89であり、これは全道路リンクの80%について累積寄与率がほぼ0.9に達していることを示す。さらに10次元増やし、特徴空間回数を30次元とした場合の累積寄与率の向上幅は0.05程度である。

表3は特徴空間予測手法(回数20次元,30次元)、日種分類手法、回帰分析について予測精度を広域的に評価した結果である。評価対象エリアは東京都全域であり、道路交通情報システムで収集されたデータのうち、予測データベース作成期間(2001年4月~2002年3月)と予測評価期間(2002年4月~7月)を通じて欠損などの異常値が5%以下である約3,200リンクについて、前節と同様にして予測誤差を算出した。評価指標は当該リンクに関する7時~19時の平均誤差率,70パーセンタイル誤差率,80パーセンタイル誤差率である。なお、回帰分析で用いた日種因子は特徴空間予測手法のそれと同じものである。

この結果からは、日種分類手法に対して広域的にも予測精度が向上していること、また、低次元化された特徴空間上で予測を行った場合でも、所要時間データ空間における通常の回帰分析と同等の予測精度が確保できることが分かる。特徴空間の回数については、20次元でも30次元でも差のないことが確認された。つ

表2 東京都内主要3,500リンクにおける累積寄与率

Table 2 Cumulative contribution ratio for 3,500 major links in Tokyo.

dimension of feature space	1	2	...	18	19	20	21	22	...	28	29	30
average	0.50	0.62	...	0.93	0.93	0.94	0.94	0.94	...	0.96	0.96	0.97
70th percentile	0.38	0.51	...	0.91	0.91	0.92	0.93	0.93	...	0.96	0.96	0.96
80th percentile	0.33	0.46	...	0.88	0.88	0.89	0.90	0.91	...	0.94	0.94	0.94
90th percentile	0.26	0.39	...	0.83	0.84	0.85	0.86	0.87	...	0.91	0.91	0.92

表3 広域予測精度とシステム運用性の評価

Table 3 Error rate for major links in Tokyo and performance of the system.

forecast method		Classification Method (three day factors)	Regression Analysis	Proposed Method (30 dimension)	Proposed Method (20 dimension)
error rate [%]	average	23.1	18.6	18.6	18.5
	70th percentile	27.6	21.0	20.9	20.9
	80th percentile	34.1	27.0	27.0	26.9
database size for 120,000 links		0.42 GB	11 GB	5.3 GB	3.5 GB
calculation amount	model determination (offline process)	very small	<RA> × M	<PCA> + <RA> × 30	<PCA> + <RA> × 20
	forecasting (online process)	very small	matrix product-sum O(D × M)	matrix product-sum O((D + M) × 30)	matrix product-sum O((D + M) × 20)

RA: Regression Analysis, PCA: Principal Component Analysis
M: Number of data for a day, D: Number of day factors

まり、累積寄与率が 0.9 程度あれば、予測対象データの低次元化による予測精度の悪化は生じないということがいえる。

表 3 には全国域のサービスに必要な予測データベースサイズならびに演算量の目安も示した。データベースサイズは、12 万リンクを対象として予測データの時間分解能を 5 分間隔とした場合の見積りであり、回帰分析と特徴空間予測手法で用いる日種因子の数は 80 種類と仮定した。この概算結果からは、特徴空間次元数を 20 次元とすることで、通常回帰分析と比べて約 1/3 程度に予測データベースを縮小できることが分かる。

予測データベースをオフラインで作成する過程では、特徴空間予測手法は主成分分析に時間を要するが、特徴空間における回帰分析の演算回数は、所要時間データ空間における通常回帰分析に比べてはるかに少ない。このことは収束演算等を要する非線形な予測モデルを用いる際に、演算時間を大きく短縮する可能性があることを示している。前章に述べたプロセス 3 により予測データを生成する処理は、日種分類手法の簡単さには及ばないものの、単純な行列積和演算で実行可能であり、動的経路探索のように高負荷な用途にも十分に耐えうる性能を示す。

5. 結 言

本研究では全国規模の所要時間予測サービスの実現を目的として、特徴空間上で予測処理を行うというアプローチにより、コンパクトな予測モデルで多様な日種因子の取扱いが可能な長期予測手法を開発した。その予測精度とシステム運用性は 4 章において実データを用いて示したとおりであり、この予測手法は実用のものでありすでに市場に提供されている。

一方、従来手法に比して予測精度や時間分解能を損なうことなく、全国域の予測システムを PC サーバで運用可能なサイズにまでスケールダウンしたとはいえ、ホームユースの PC で軽快に動かすにはまだまだ大きい。目指すのは業務用途からコンシューマ向けアプリケーションまで、スケールを問わずに適用できる技術であり、予測プロセスのさらなる効率向上のため、現在も継続して研究を進めている。これが実現すれば、たとえば予測に必要なシステムは端末に置き、通信負荷の低いリアルタイム情報の配信のみをオンラインで受けて、端末側でリアルタイム情報と予測情報をマージする、そのような使い方もできるようになる。

どこへ出かけるときも知りたい情報が手に入る、それが理想とするサービス環境だが、その実現までには

他にも解決すべき課題がある。その 1 つは既存の道路交通情報システムでカバーされていない路線の扱いである。この課題について筆者らのグループは、交通工学の見地から交通情報を空間的に推定補間する技術をカーナビ向けに実用化し⁹⁾、また、プローブカーの必要台数と情報品質について定量的な見通しを立てるなどの取組みを進めており¹⁰⁾、これら関連技術を組み合わせることで、より有用な交通情報サービスの普及を進めていく考えである。

参 考 文 献

- 1) 伏木 匠ほか：プローブカーを利用した交通情報予測方式の検討，情報処理学会論文誌，Vol.43, No.12, pp.3801-3808 (2002).
- 2) Inoue, T., et al.: A Report of Traffic Safety Aptitude Test Using the Fleet Management System, *Proc. 10th World Congress on ITS*, Madrid, CD-ROM (2003).
- 3) Rice, J., et al.: A Simple and Effective Method for Predicting Travel Times on Freeways, *Proc. IEEE 4th International Conference on ITS*, pp.227-232 (2001).
- 4) Innamaa, S.: Short-Term Prediction of Highway Travel Time Using MLP-neural Networks, *Proc. 8th World Congress on ITS*, Sydney, CD-ROM (2001).
- 5) 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).
- 6) 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).
- 7) 割田 博ほか：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).
- 8) 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).
- 9) Yamane, K., et al.: Statistical Traffic Information for Navigation System, *Proc. 11th World Congress on ITS*, Nagoya, CD-ROM (2004).
- 10) 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).

(平成 16 年 4 月 5 日受付)

(平成 16 年 10 月 4 日採録)



熊谷 正俊 (正会員)

1975年生。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)。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、日本ロボット学会会員。1998年度日本機械学会東北支部独創研究学生賞受賞。



佐野 豊

1969年生。1994年3月東北大学工学研究科電気および通信工学専攻博士課程前期修了。同年4月(株)日立製作所入社。以来、同社情報制御システム事業部において、交通情報提供、交通管制、ETC、画像処理等ITS関連設備およびシステムの開発、SE、エンジニアリングに従事。電気学会会員。1999年第48回電機工業技術功績者表彰(進歩賞)受賞。1級電気工事施工管理技士。



伏木 匠 (正会員)

1973年生。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16年度電機工業技術功績者奨励賞受賞。



鈴木 研二

1976年生。2002年3月北海道大学大学院工学研究科電子情報工学専攻修士課程修了。同年4月(株)日立製作所入社。以来、同社情報制御システム事業部において、交通情報提供、画像処理等ITS関連設備およびシステムの開発、SE、エンジニアリングに従事。



横田 孝義 (正会員)

1956年生。1984年3月東京工業大学総合理工学研究課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。電気学会会員。