



# データマイニングによる 交通情報予測システムの開発

森田 哲郎

Traffic Condition Forecasting System Based on Probe Car Data with Data Mining Method —— by Tetsuo Morita —— A multi-agent system which can forecast traffic conditions based on probe car data with data mining method has been developed. This system consists of estimation and learning agents assigned to road links. Estimation agents renew normalized congestion level for each road link, while learning agents renew weight values for estimation. The weight values are calculated by multivariate analysis. Estimation and learning agents provide calculated results alternately to improve the accuracy of forecast. To forecast future traffic conditions, it is essential to interpolate the present traffic condition data. In this study, the results which were forecasted one hour in advance were used as the interpolation results for the following hour. Based on the probe car data collected from 2007/11/01 to 2008/10/31, the mean square error of an-hour-ahead forecasting velocity was calculated to be  $1.9679 \times 10^{-2}$  and the standard deviation of the forecasting velocity error was 14.03 (km/h).

Keywords: data mining, probe car, multivariate analysis, coefficient of determination, mean square error

## 1. 緒 言

交通渋滞の発生を事前に予測することができれば、より少ない時間で目的地まで到着することが可能になり、燃料消費や二酸化炭素の排出量を削減することができる。交通情報サービスは、上記の問題を解決する有効な手段であると考えられる。交通情報サービスの役割としては、数時間後に道路が混雑し始めるといった「予測（時間）情報の提供」と渋滞表示マップのような「空間情報の提供」に分類できる。

ドライバに渋滞に関する空間情報を提供する方法としては、道路に感知器を設置し、その感知器から得られた速度（混雑状況）をドライバに提供するシステムとしてVICS（Vehicle Information and Communication System）があり、さらにVICSを用いて道路の混雑状況を予測するシステム<sup>(1)</sup>が提案されている。VICSは、道路に感知器を設置する必要があり、日本全国に感知器を設置するためには、膨大な設備投資が必要となる。この問題を解決する方法として、注目を集めているのがプローブカー・システムである。プローブカー・システムでは、車両に設置されたGPS（Global Positioning System）等のデータから、道路リンクの旅行時間を得ることができる。プローブカーを用いたシステムは、道路に感知器を設置する必要がないため、インフラ整備に頼らずに交通情報を提供することが可能になる。しかし現状では、プローブカーの台数は非常に少なく、プローブカー・データ（PCD）のみで必要とする道路リンクの旅行時間を知ることはできない。

PCDの欠落を補うための方法として、フェロモンの放

出・拡散・蒸発を用いた渋滞予測システム<sup>(2)~(4)</sup>が提案されている。これは、予測という「時間情報の提供」と補完という「空間情報の提供」の両方に用いることが可能である。車速に応じて一定量のフェロモンが放出され、一定の割合で拡散・蒸発を繰り返すというメカニズムによって、周辺道路リンクの旅行時間を推定することが可能になる。しかし、フェロモンの放出・拡散・蒸発に関するパラメータ設定は、人が経験に基づいて行うため、客観性に欠けるという欠点がある。

プローブカーからの欠落したリンク旅行時間を埋める空間補完技術として、欠損値付き主成分分析<sup>(5)、(6)</sup>という手法が提案されている。欠損値のない主成分分析は、データマイニングで一般的に用いられる手法で、道路リンク間の相関を求め、その相関の高さから他の道路リンクの旅行時間を推定する方法である。しかし、複数の道路リンクを同一時刻にプローブカーが走行する確率が極めて低いために、欠損値付き主成分分析を用いる必要がある。またこの方法では、特定の領域に存在する全ての道路リンクの相関を取るため、膨大なCPU時間を必要とする。

上記の問題を解決するために、(1) フェロモンモデルで客観的に決定できない放出・拡散・蒸発に関するパラメータを「学習」によって決定する、(2) 欠損値付き主成分分析で問題となる欠損値を、周辺道路リンクから得られるPCDを用いて「推定」する交通情報の空間補完システム<sup>(7)~(9)</sup>を提案した。この方法では、指定された領域に含まれる道路リンクに配置された推定エージェントが、PCDの得られて

いない道路リンクの速度を推定し、学習エージェントがデータマイニングの手法<sup>(10)</sup>を用いて、推定の精度を高めている。

同様の原理を利用して、プローブカーから得られる速度データを基に交通情報の予測を行うシステムを開発した。交通情報の空間補完システムでは、推定エージェントが現在の周辺道路リンクの速度を基に現在の道路リンクの速度を推定していたが、交通情報予測システムでは、過去ならびに現在の周辺道路リンクの速度を参照することによって、未来の道路リンクの速度を予測するシステムである。今回の予測システムにおいて、推定エージェントと学習エージェントを交互に動作させることによって、予測精度が改善することを確認した。

## 2. 交通情報予測システム

**2-1 交通情報予測システム** 交通情報予測システムで、渋滞の度合いを示すために（規格化された）渋滞度を用いる。空間補完システムでは、規格化速度を用いたが、（規格化）渋滞度は、**図1**に示すように、規格化速度と同様の概念である<sup>(7)~(9)</sup>。渋滞度は、0から1に規格化され、道路リンクを通過する車両の速度が速くなるほど、渋滞度は0に近づく。まったく車両が動かない状態では、渋滞度は1になる。

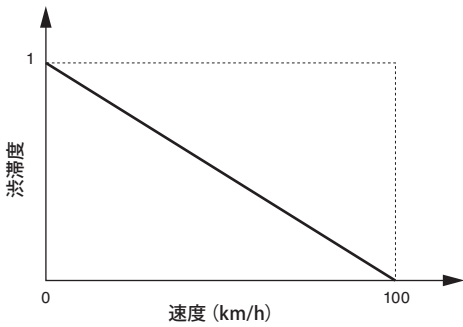


図1 渋滞度の定義

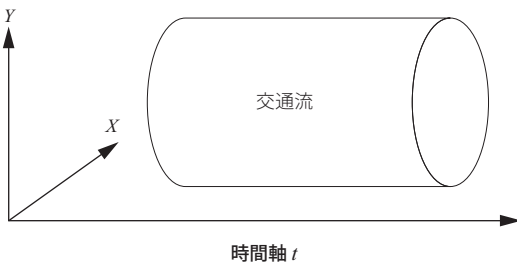


図2 交通流のイメージ

道路の混雑状況は、時間と空間の関数で記述され、**図2**に示すように、交通流連続体として捉えることができる。特定の道路リンクの渋滞度は、その道路リンク自体の過去の渋滞度と周辺道路リンクの渋滞度で、決まると考えられる。従来は、1つの道路リンクのみに着目し、一定の周期で道路リンクの渋滞度が変化すると仮定して、未来の渋滞度を予測していた。例えば、過去数週間の月曜日午前9時の渋滞度の平均値を、次の月曜日の午前9時の渋滞度として予測する方法である。つまり**図3**に示すように、交通流連続体を時間軸が含まれるように切断して、時間軸上で渋滞度を予測している。

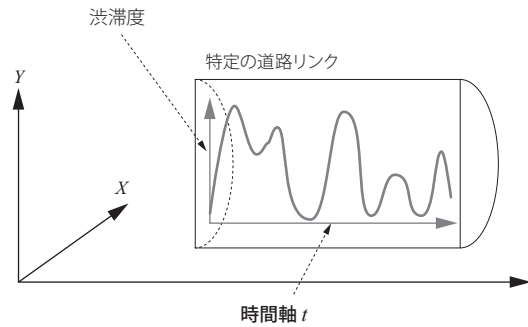


図3 渋滞度の時間軸上での変化

周期関数を数学的に記述する方法として、フーリエ級数展開が知られている。すべての周期関数は、**式(1)**に示されるように、正弦波 (sin) と余弦波 (cos) の級数展開で得ることができる。 $\omega$ は角周波数であり、 $a_h, b_h$ は係数である。**式(1)**では、無限個の $a_h, b_h$ を決定する必要があるが、実際の交通情報システムでは、一定の時間毎（離散的）に道路リンクからPCDが得られるため、有限個で追従可能である。周期を1週間、PCDの得られる時間間隔を15分間（1時間に4回）とすると、 $h$ は1から336（＝7日×24時間×4÷2）となる。周期が長くなると（相対的に）高い周波数成分のsin波、cos波まで計算する必要があり、数多くの係数 $a_h, b_h$ を決める必要がある。このように時間軸が含まれるように交通流連続体を切断すると、渋滞度は時間軸上で非線形関数となる。

$$g_0(t) = \frac{a_0}{2} + \sum_{h=1}^{\infty} \{a_h \cos(h\omega t) + b_h \sin(h\omega t)\} \quad \dots\dots (1)$$

**図4**は、交通流連続体を空間軸が含まれるように切断したイメージである。空間軸が含まれるように切断すると、切断面での渋滞度は線形関数で記述できると考えられる。渋滞度を線形問題に落とし込むことができれば、多変量解析（データマイニング）の手法を適用することができる。

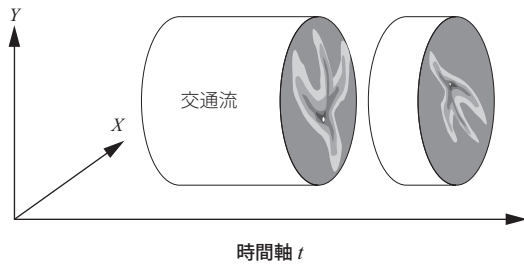


図4 渋滞度の空間軸上での変化

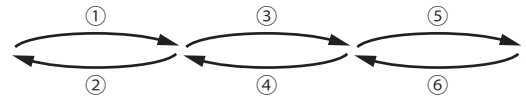


図7 道路リンクの接続トポロジ

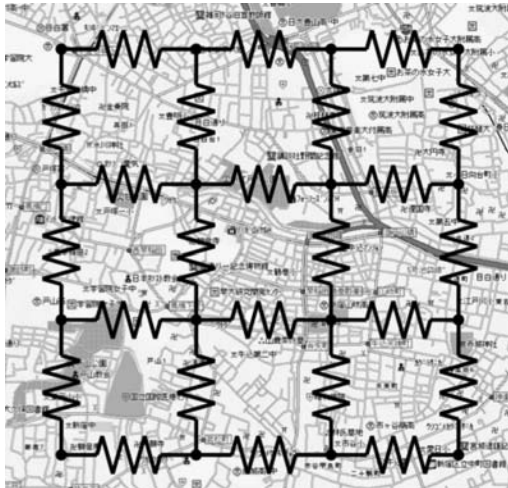


図5 双対グラフを用いた道路ネットワーク

つまり、知りたい道路リンクの渋滞度は、隣接道路リンクの渋滞度の線形和によって、表すことができる。

空間軸が含まれるように交通流連続体を切断した場合、道路リンクは図5の線形抵抗ネットワークで表されると考えられる。この図からは、抵抗が道路リンクに相当し、接点が交差点であるという印象を受けるが、実際には双対グラフで表されているため、逆の関係になる。

図6に、実際の道路形状を示す。1本の対面通行道路は、信号等によって複数の道路に分割され、分割された単位が道路リンクと呼ばれている。図6の例では、反対向きの車線を含めて6つの道路リンクに分割され、それぞれに①か

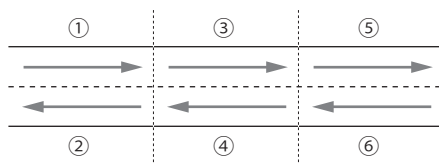


図6 実際の道路形状

ら⑥の番号が振られている。

この道路リンクを接続トポロジで示したのが、図7である。各道路リンクは、反対方向の道路リンクと対になるため、両端で接続している。

図8は、図7の接続トポロジを双対グラフで表したものである。図8では、道路リンクが接点として表され、道路リンク③は、他の全ての道路リンクと接続しているため、5つの辺(エッジ)で接続されている。図8で $V_1$ から $V_6$ は、各道路リンクの渋滞度で、回路理論的に考えると電圧に対応する値である。道路リンク③は、他の全ての道路リンク

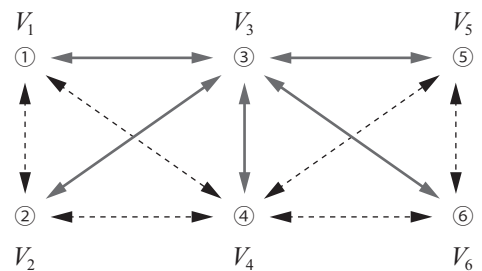


図8 道路リンクの接続関係(双対グラフ)

と抵抗で接続されている。

道路リンク①から道路リンク③に流入する渋滞度(電流)は、道路リンクの渋滞度(電圧)差である $V_1 - V_3$ とコンダクタンスに相当する重み $w'_{(1 \rightarrow 3)}$ との積によって与えられる。つまり、隣接する道路リンク①から電流に相当する渋滞度が道路リンク③に流れ込み、渋滞が引き起こされる。

電気回路の基本法則であるキルヒホッフの第一法則を、道路リンク③に対して適用すると、道路リンク③に流れ込む各道路リンクからの電流の合計値は0となり、式(2)が得られる。重み値 $w'_{(1 \rightarrow 3)}$ は、道路リンク①から道路リンク③へのコンダクタンスを示している。

$$w'_{(1 \rightarrow 3)}(V_1 - V_3) + w'_{(2 \rightarrow 3)}(V_2 - V_3) + w'_{(4 \rightarrow 3)}(V_4 - V_3) + w'_{(5 \rightarrow 3)}(V_5 - V_3) + w'_{(6 \rightarrow 3)}(V_6 - V_3) = 0 \quad \dots\dots\dots (2)$$

このシステムで扱うコンダクタンスは異方性があり、道路リンク③から見たコンダクタンス $w'_{(1 \rightarrow 3)}$ と道路リンク①

から見たコンダクタンス  $w'_{(3 \rightarrow 1)}$  は、異なる値を取る。道路リンク間のコンダクタンスである  $w'_{(1 \rightarrow 3)}$  等の値が小さい場合は、隣接道路リンクから渋滞の影響を受けにくい（渋滞流が流れ込みにくい）道路となる。また、道路リンク③と④は対面する道路であり、道路リンク④から③へ流入する車両はほとんど無いと考えられるが、脇見渋滞が発生する可能性もあるため、渋滞流が流れ込むと考える。

式(2)を  $V_3$  に関してまとめると、式(3)になる。ここで各係数は、式(4)(5)で与えられる。

$$V_3 = w_{(1 \rightarrow 3)}V_1 + w_{(2 \rightarrow 3)}V_2 + w_{(4 \rightarrow 3)}V_4 + w_{(5 \rightarrow 3)}V_5 + w_{(6 \rightarrow 3)}V_6 \quad \dots (3)$$

$$w_{(1 \rightarrow 3)} = \frac{w'_{(1 \rightarrow 3)}}{5\bar{w}_{(3)}}, w_{(2 \rightarrow 3)} = \frac{w'_{(2 \rightarrow 3)}}{5\bar{w}_{(3)}}, w_{(4 \rightarrow 3)} = \frac{w'_{(4 \rightarrow 3)}}{5\bar{w}_{(3)}},$$

$$w_{(5 \rightarrow 3)} = \frac{w'_{(5 \rightarrow 3)}}{5\bar{w}_{(3)}}, w_{(6 \rightarrow 3)} = \frac{w'_{(6 \rightarrow 3)}}{5\bar{w}_{(3)}} \quad \dots (4)$$

$$5\bar{w}_{(3)} = w'_{(1 \rightarrow 3)} + w'_{(2 \rightarrow 3)} + w'_{(4 \rightarrow 3)} + w'_{(5 \rightarrow 3)} + w'_{(6 \rightarrow 3)} \quad \dots (5)$$

キルヒホッフの第一法則を用いることによって、道路リンク③の電圧は、隣接する道路リンク①②④⑤⑥の電圧の線形和で記述できる。別の言い方をすると、線形抵抗による隣接道路リンクからの渋滞流の流れ込みによって、対象となる道路リンクの渋滞度が決まる。

図9は、(一方通行の)道路リンクの接続例である。 $g_1(t)$ ,  $g_2(t)$ ,  $g_3(t)$ ,  $g_4(t)$ は、対応する各道路リンクの渋滞度を時刻  $t$  で表す関数である。図9のように道路リンクが接続されている場合は、 $g_0(t)$  に対応する道路リンクの渋滞度は、式(6)のように表される。つまり、 $g_1(t)$ ,  $g_2(t)$ ,  $g_3(t)$ ,  $g_4(t)$  に対応する隣接道路リンクは時刻  $t$  の非線形関数であるが、 $g_0(t)$  と  $g_1(t)$ ,  $g_2(t)$ ,  $g_3(t)$ ,  $g_4(t)$  との関係式は、線形関数で表すことができ、多変量解析によるデータマイニングの手法を適用することが可能になる。

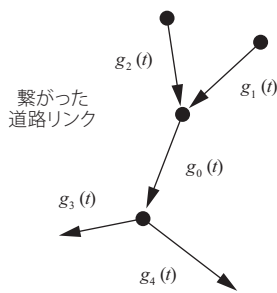


図9 道路リンクの接続例

$$g_0(t) = w_{(1 \rightarrow 0)}g_1(t) + w_{(2 \rightarrow 0)}g_2(t) + w_{(3 \rightarrow 0)}g_3(t) + w_{(4 \rightarrow 0)}g_4(t) \quad \dots (6)$$

式(6)の場合は、項数が隣接道路リンク数によって決まるため、式(1)より項数が少ないことが分かる。つまり  $g_0(t)$  を式で表現する際に、類似性の少ない sin 関数や cos 関数を利用するのではなく、類似している隣接道路リンク波形の線形和で  $g_0(t)$  を得ることができるため、少ない項数で  $g_0(t)$  を記述できる。別の見方をすると、空間軸が含まれるように交通流連続体を切断すると、時刻  $t$  が暗黙変数となり、時間項が消去されたように見える。

一方、学習対象となる道路リンクと隣接道路リンクの渋滞度が酷似している場合、重回帰分析では「多重共線性」という問題が発生することが知られている。この問題を解決するためには、各道路リンク間の相関係数を調べ、例えば相関係数が0.8以上の隣接道路リンクを参照しないことによって、予測システムの精度を改善することができる。

今回使用するPCDは、対象領域に存在する道路リンクの数パーセントから得られるのみである。図8の例では、道路リンク③を推定する場合、道路リンク①②④⑤⑥の渋滞度を知る必要があるが、どの道路リンクからもPCDを得られない可能性がある。交通情報補完システムでは、推定エージェントが、PCDの得られた道路リンクの渋滞度（規格化速度）を基に、PCDの得られていない道路リンクの渋滞度（規格化速度）を推定していた。しかし交通情報予測システムの場合は、(例えば)1時間先の交通状況を予測する場合には、未来のPCDを得ることはできない。従って予測システムでは、現在と過去の渋滞度の関係をデータマイニング（重回帰分析）し、その結果を基に未来の渋滞度を予測する。つまり、渋滞流の流れ込むという点では予測システムは空間補完システムと同様であるが、予測システムでは渋滞流が過去ならびに現在から、未来へ流れ込む点が異なっている。また学習エージェントは、予測された未来の時刻が現在時刻になった（1時間先を予測して、1時間経過した）時点で、PCDが得られた道路リンクに対して、学習（データマイニング）を行う。

図10に、交通情報予測システムの構成を示す。各道路リンクに、推定エージェントと学習エージェントの2つを配置する。各道路リンクの渋滞度は推定エージェントに

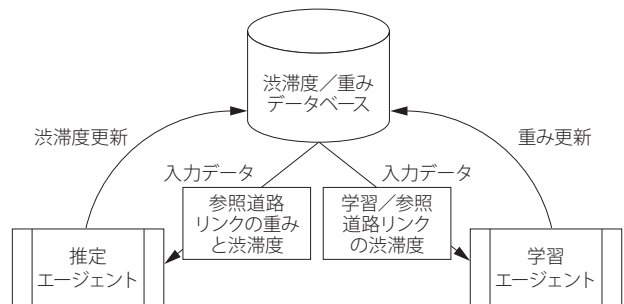


図10 交通情報予測システムの構成

よって更新され、計算に必要な重み値は学習エージェント(重回帰分析)<sup>(10)</sup>によって更新される。推定された渋滞度と学習によって得られた重み値は、道路リンクの渋滞度/重みデータベースに保存される。交通情報予測システムはPCDを読み込み、推定エージェントと学習エージェントを交互に動作させることによって、予測精度を高めることが可能となる。

**2-2 推定エージェント** 推定エージェントは、予測対象となる道路リンク(以下、予測道路リンク)に隣接する道路リンク(以下、参照道路リンク)の渋滞度と学習エージェントが算出した最新の重み値から、予測道路リンクの $f(f \geq 1)$ タイムスロット先の渋滞度を算出する。推定/学習時に(同一路径リンクの)参照するタイムスロット数を $g(g \geq 1)$ とする。予測道路リンク $i$ に対する参照道路リンク数 $n^{(i)}$ と、現在時刻 $t$ から $t-g+1$ までのタイムスロット数 $g$ との積である $gn^{(i)}$ 個の渋滞度に、定数1を追加した渋滞度ベクトルを ${}^tV_{gn+1}^{(i)}$ 、現在時刻 $t$ の重みベクトルを ${}^tW^{(i)}$ とすると、 $f$ タイムスロット先である時刻 $t+f$ の予測渋滞度 ${}^{t+f}\tilde{E}^{(i)}$ は、式(7)のように算出される。予測道路リンク $i$ に対する参照道路リンク数は $n^{(i)}$ 個であり、道路リンク毎に参照道路リンク数は異なるが、誤解を生じない場合は $n^{(i)}$ を単に $n$ と表記する。

$${}^{t+f}\tilde{E}^{(i)} = {}^tV_{gn+1}^{(i)} \cdot {}^tW^{(i)} \quad \dots\dots\dots (7)$$

$${}^tV_{gn+1}^{(i)} = (1 \ {}^tV_n^{(i)} \ \dots \ {}^{t-g+1}V_n^{(i)}) \quad \dots\dots\dots (8)$$

$${}^tV_n^{(i)} = ({}^tV_1^{(i)} \ {}^tV_2^{(i)} \ \dots \ {}^tV_n^{(i)}) \quad \dots\dots\dots (9)$$

$${}^tW^{(i)} = ({}^tW_0^{(i)} \ {}^tW_1^{(i)} \ \dots \ {}^tW_{gn}^{(i)})^T \quad \dots\dots\dots (10)$$

参照道路リンクの渋滞度は、他の推定エージェントが算出した渋滞度、もしくはその時刻に得られたPCDによる渋滞度が用いられる。全ての道路リンクの渋滞度の初期値は0であり、重みの初期値は $w_0^{(i)} = 0$ ,  $w_1^{(i)} + \dots + w_{gn}^{(i)} = 1$ となる正の乱数値を割り当てる。 $w_0^{(i)}$ は、閾値と呼ばれることもある。

**2-3 学習エージェント** 学習エージェントは、学習対象となる道路リンク(以下、学習道路リンク)から得られた(PCDによる)渋滞度と、学習道路リンクに隣接する道路リンク(参照道路リンク)の渋滞度から、参照道路リンクの重みベクトル ${}^tW^{(i)}$ を算出する。推定の場合と同様に、学習道路リンク $i$ に対する参照道路リンク数を、 $n$ と表記する。学習道路リンク $i$ から得られるPCD数 $m^{(i)}$ に関しても $m$ と表記し、重みベクトル ${}^tW^{(i)}$ の添字 $t$ ならびに $(i)$ も省略する。

学習道路リンク $i$ から時刻 $t(1), \dots, t(m)$ に得られた $m$ 個の(PCDによる)渋滞度をプローブベクトル $\mathbf{P}$ とする。誤解を生じない場合は、添字 $(i)$ を省略する。

$$\mathbf{P} = (P_1 \ P_2 \ \dots \ P_m)^T \quad \dots\dots\dots (11)$$

PCDが得られた時刻 $t(j)$ の $f$ タイムスロット前の参照道路リンクの渋滞度ベクトル ${}^{t(j)-f}V_n$ を、式(12)のように表す。参照道路リンクの渋滞度は、推定エージェントが算出した渋滞度もしくはその時刻に得られた(PCDによる)渋滞度が用いられる。

$${}^{t(j)-f}V_n = ({}^{t(j)-f}V_1 \ {}^{t(j)-f}V_2 \ \dots \ {}^{t(j)-f}V_n) \quad \dots\dots\dots (12)$$

渋滞度ベクトル ${}^{t(j)-f}V_n$ を行ベクトルとして持つ渋滞度マトリックス ${}^{-f}V_{m \times n}$ は、下記のように表される。

$${}^{-f}V_{m \times n} = \begin{pmatrix} {}^{t(1)-f}V_n \\ {}^{t(2)-f}V_n \\ \vdots \\ {}^{t(m)-f}V_n \end{pmatrix} \quad \dots\dots\dots (13)$$

交通情報予測システムは、学習道路リンク $i$ の現在時刻 $t$ から $f$ タイムスロット先の渋滞度を予測する。 $f$ タイムスロット先を予測するために、学習時には(PCDによる)渋滞度とPCDが得られた時刻から $f$ タイムスロット前 $(-f)$ から $f+g-1$ タイムスロット前 $(-f-g+1)$ までの渋滞度との関係を学習する。従って、学習に使用する渋滞度マトリックスは、式(14)のようになる。 $\mathbf{I}_m$ は、 $m$ 個の1を要素として持つ列ベクトルである。

$${}^{-f}V_{m \times (gn+1)} = (\mathbf{I}_m \ {}^{-f}V_{m \times n} \ \dots \ {}^{-f-g+1}V_{m \times n}) \quad \dots\dots\dots (14)$$

学習道路リンクから得られたプローブベクトル $\mathbf{P}$ を、渋滞度マトリックス ${}^{-f}V_{m \times (gn+1)}$ と重みベクトル $\mathbf{w}$ で表すと、式(15)に示される方程式数 $m$ 、未知数 $gn+1$ の連立方程式を得る。

$$\mathbf{P} = {}^{-f}V_{m \times (gn+1)} \cdot \mathbf{w} \quad \dots\dots\dots (15)$$

方程式の数 $m$ が未知数の数 $gn+1$ を下回った場合( $m < gn+1$ )、解である重みベクトル $\mathbf{w}$ が定まらない。 ${}^{-f}V_{m \times (gn+1)}$ の階数が $gn+1$ の場合は、連立方程式の解が得られる。一次独立な方程式の数が未知数の数 $gn+1$ を上回ると、解である重みベクトルは存在しない。この場合は、残差平方和を最小にする方法、つまり最小自乗法を用いる。参照道路リンクの渋滞度マトリックス ${}^{-f}V_{m \times (gn+1)}$ と、重みベクトル $\mathbf{w}$ から算出される渋滞度の予測ベクトル $\mathbf{E}$ は、式(16)で表される。

$$\mathbf{E} = {}^{-f}V_{m \times (gn+1)} \cdot \mathbf{w} \quad \dots\dots\dots (16)$$

$$\mathbf{E} = (E_1 \ E_2 \ \dots \ E_m)^T \quad \dots\dots\dots (17)$$

プローブベクトル  $P$  と予測ベクトル  $E$  の  $k$  番目の成分の差、つまり残差を  $\varepsilon_k$  とすると、残差平方和  $Q$  は下記の式で与えられる。

$$Q = \sum_{k=1}^m \varepsilon_k^2 = \sum_{k=1}^m (P_k - E_k)^2 \quad \dots\dots\dots (18)$$

残差平方和が最小になるのは、式 (18) を  $gn+1$  個の重み  $w_0, w_1, \dots, w_{gn}$  で偏微分した式が、下記の条件を満たす時である。

$$\frac{\partial Q}{\partial w_u} = 0 \quad (u = 0, 1, \dots, gn) \quad \dots\dots\dots (19)$$

### 3. 交通情報予測システムの評価

交通情報予測システムを動作させる際には、現在時刻の渋滞度を知る必要があるが、今回の評価には空間補完を用いずに、予測システムのみを用いた。予測システムが1時間先の渋滞度を予測し、1時間経過後に、その予測値を空間補完された現在の渋滞度として利用している。さらに、その空間補完された渋滞度を基に、1時間先の道路リンクの渋滞度を予測している。

使用したデータは、名古屋のタクシープローブであり、評価対象領域は、名古屋駅を含む約10km四方である。タクシーの台数は約1,200台で、総道路リンク数は1,128である。評価期間は2007/11/1～2008/10/31であり、PCDは15分間に1回得られるため、1タイムスロットは15分間となる。また1時間先の交通情報を、現在と1つ前のタイムスロットから予測するため、 $f=4, g=2$  となる。PCDは、平均して5%の道路リンクから得られるため、1つの道路リンクに着目すると、5時間 (= 15分間÷0.05) に1回PCDが得られることになる。

**3-1 決定係数ならびに残差自乗平均による評価** 交通情報予測システムは、同一状況下で、対象とする道路リンクを通過する車両の平均速度を求めることを目標としている。各道路リンクの決定係数  $R$  ならびに残差自乗平均  $Q/m$  を用いると、その性能を評価することが可能になる。重回帰分析で用いられる学習道路リンクの決定係数  $R$  は、式 (20) で与えられる。

$$R = \frac{\sum_{j=1}^m (E_j - \bar{E})^2}{\sum_{j=1}^m (P_j - \bar{P})^2} \quad \dots\dots\dots (20)$$

$$\bar{E} = \frac{1}{m} \sum_{j=1}^m E_j \quad \dots\dots\dots (21)$$

$$\bar{P} = \frac{1}{m} \sum_{j=1}^m P_j \quad \dots\dots\dots (22)$$

図11に、1年間に2,000～2,500個程度のPCDが得ら

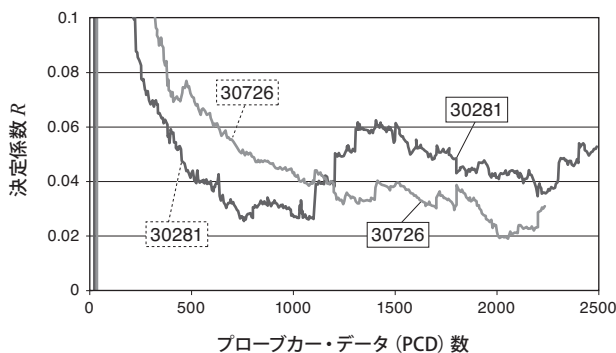


図11 決定係数の変化  
(PCDの最低保存数：1,000, PCDの最低保存日数：30日)

れる道路リンクを選択し、決定係数の変化を調べた結果を示す。横軸はPCD数であり、縦軸は決定係数である。道路リンク番号は、5桁の数字で表し、5桁目が1の場合は都市間高速道路、2の場合は都市内高速道路、3の場合は一般道であり、1桁目から4桁目は、割り当てられた道路リンク番号を示す。従って、図11で選ばれた道路リンクは、一般道であることが分かる。

学習結果である重み値を1回も得られない道路リンク、つまり学習を開始していない道路リンクに対しては、決定係数を0として扱う。式 (15) の連立方程式が解けた場合、決定係数は1になる。図11では、グラフの上限値が0.1であるため、はみ出している。さらにPCD数が増加すると、式 (19) の連立方程式を解くことになる。

空間補完の結果<sup>(7)～(9)</sup>では、PCD数が増えるにつれて決定係数が増加する傾向があったが、今回は決定係数が増加していない事が分かる。この原因は、明示的に空間補完を行わず、1時間先の予測結果が現在時刻になった時点で、その予測結果を空間補完による渋滞度として利用した点にあると思われる。

PCD数が1100個を超えた段階で、決定係数が急激に増加しているのは、保存してあるPCDが1100個に達すると、古い100個のPCDを廃棄するためである。この後、PCDの保存数が1100個を越えないように、100個ずつ削除されている。ちなみに、PCDの最低保存日数は30日であり、PCDの保存数が1100個を越えても、30日以上経過していないPCDは廃棄されない。横軸のPCD数が1100個以上の範囲では、決定係数は概ね0.02から0.06の間にあり、ほぼ一定の値を示している。

重回帰分析では、残差自乗平均を最小にする重み値を決定する。決定係数と並んで、残差自乗平均も交通情報予測システムを評価する重要なパラメータである。図12に、残差自乗平均  $Q/m$  の変化を示す。横軸はPCD数であり、縦軸は残差自乗平均である。残差自乗平均は、PCDが得られない段階では1に設定し、プローブが得られた段階で初

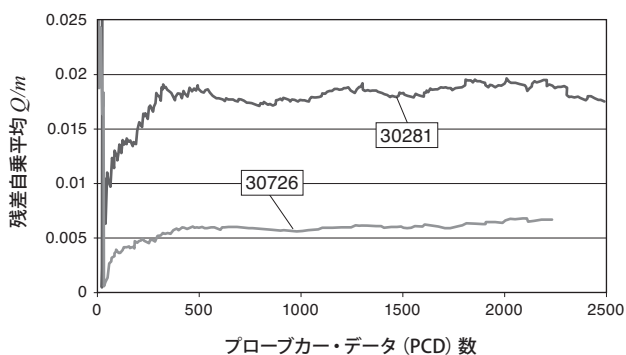


図12 残差自乗平均の変化  
(PCDの最低保存数：1,000, PCDの最低保存日数：30日)

期重みで記述された超平面との残差自乗平均を用い、さらに学習を開始した段階(式(15)が解けた時点)で0になる。その後、急激に増加した後、ほぼ一定の値を示している。ちなみに、分散に相当する残差自乗平均が0.01の場合、渋滞度の標準偏差は0.1となり、速度誤差の標準偏差は図1の渋滞度の定義から10(km/h)となる。

空間補完の結果(7)~(9)では、PCD数が増えるにつれて残差自乗平均が減少する傾向があったが、予測システムでは、ほぼ一定の値を示している。この原因も、明示的に空間補完を行わなかった点にあると思われる。

3-2 評価領域全体での評価 評価対象領域に存在するN個の道路リンクに対する決定係数の平均値 $R_{ave}$ ならびに残差自乗平均の平均値 $Q_{ave}$ は、式(23)(24)で与えられる。式(23)(24)では、道路リンク番号を区別するために、 $R^{(i)}$ 、 $Q^{(i)}$ 等の表記を用いる。

$$R_{ave} = \frac{1}{N} \sum_{i=1}^N R^{(i)} \quad \dots\dots\dots (23)$$

$$Q_{ave} = \frac{1}{N} \sum_{i=1}^N \frac{Q^{(i)}}{m^{(i)}} \quad \dots\dots\dots (24)$$

図13に、評価対象領域に含まれる全道路リンクの決定係数の平均値 $R_{ave}$ ならびに残差自乗平均の平均値 $Q_{ave}$ の変化を示す。横軸は日付で、縦軸は決定係数ならびに残差自乗平均の平均値である。対象期間は、2007/11/1~2008/10/31の12ヶ月間であり、学習が進むにつれて、 $R_{ave}$ ならびに $Q_{ave}$ は減少していることが分かる。 $Q_{ave}$ の値が急激に減少しているのは、正月、建国記念日、ゴールデンウィークで、これまで1回もPCDを得られなかった道路リンクから、PCDを得たためである。この時期に名古屋のタクシーは、通常通過しない道路リンクを走行すると思われる。

空間補完の結果(7)~(9)では、決定係数は徐々に増加していたが、今回の予測結果では、減少していることが分かる。この原因も、明示的に空間補完を行わず、1時間先の予測結果が現在時刻になった時点で、その予測結果を空間補完

による渋滞度として利用した点にあると思われる。

評価の終了した2008/10/31の時点で、残差自乗平均の平均値は、 $6.755 \times 10^{-2}$ であり、この時点でPCDの得られない道路リンクは54個である。PCDの得られない道路リンクの残差自乗平均は1として取り扱われている。PCDの得られない道路リンクは補完することができないため、補完可能な道路リンクの残差自乗平均の平均値は、 $1.9679 \times 10^{-2}$  ( $= 6.7551 \times 10^{-2} - 54/1128$ )となり、渋滞度での予測誤差の標準偏差は、0.1403となる。渋滞度の定義(図1参照)に基づいて、この標準偏差を速度に変換すると予測誤差の標準偏差は14.03(km/h)となる。今後は、現在時刻の渋滞度を推定する場合に、空間補完を用いることによって予測精度の向上を図る予定である。

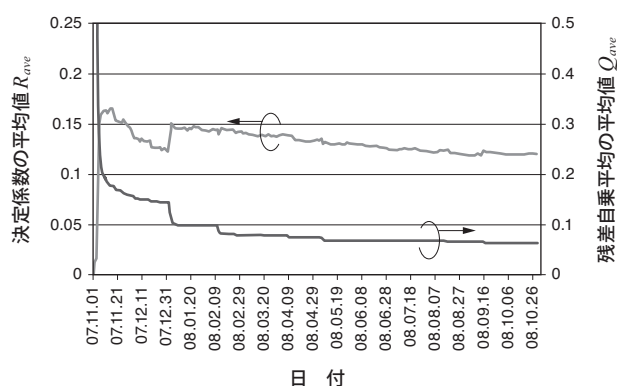


図13 評価領域全体での決定係数ならびに残差自乗平均の平均値

#### 4. 結 言

データマイニングの手法(重回帰分析)を用いて、交通情報の予測を行うシステムを開発した。推定エージェントは、参照道路リンクから得られた渋滞度を用いて予測道路リンクの渋滞度を算出し、予測に必要な重み値は学習エージェントが算出する。1時間先を予測するためには、PCDの得られない道路リンクの渋滞度を空間補完する必要があるが、1時間先の予測値が現在時刻になった際の渋滞度を補完結果として用いた。交通情報の予測システムの評価を行い、予測誤差の標準偏差14.03(km/h)を得た。

・VICSは、(財)道路交通情報通信センターの登録商標です。

※1 VICS

渋滞や交通規制などの道路交通情報をリアルタイムに送信し、カーナビゲーションなどの車載機に文字・図形で表示する情報通信システム。道路にカメラ等のセンサを設置する必要がある。

※2 プロブカー・システム

GPS等のセンサーが装着された車からのデータを収集し、ナビゲーションシステムなど走行中の車へ、渋滞など交通情報の提供を行うシステム。道路にセンサーを設置するのではなく、車載機からのデータを基に情報提供を行う。フローティングカー・システムと呼ばれることもある。

参考文献

- (1) 平松綾子、能勢和夫、天目健二、森田剛史、「状態方程式に基づく市街地幹線道路の旅行時間予測」、電気学会論文誌C（電子・情報・システム部門誌）、Vol.128, No.2, pp.269-277（2008）
- (2) Y. Ando, Y. Fukazawa, O. Masutani, H. Iwasaki, S. Honiden, "Performance of Pheromone Model for Predicting Traffic Congestion", Proc. of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 73-80（2006）
- (3) S. Kurihara, H. Tamaki, M. Numao, K. Kagawa, J. Yano, T. Morita, "Traffic Congestion Forecasting based on Pheromone Communication Model for Intelligent Transport Systems", IEEE Congress on Evolutionary Computation, Trondheim, pp. 2879-2884（May 2009）
- (4) S. Kurihara, H. Tamaki, M. Numao, K. Kagawa, J. Yano, T. Morita, "Traffic Congestion Forecasting based on Ant Model for Intelligent Transport Systems.", The 3rd International Workshop on Emergent Intelligence in Networked Agents (WEIN 2009) (Workshop at AAMAS 2009), Budapest, pp. 42-47（May 2009）
- (5) M. Kumagai, T. Fushiki, K. Kimita, T. Yokota, "Long-range Traffic Condition Forecast using Feature Space Projection Method", Proc. of 11th World Congress of ITS, Nagoya, CD-ROM（Oct. 2004）
- (6) M. Kumagai, T. Fushiki, K. Kimita, T. Yokota, "Spatial Interpolation of Real-time Floating Car Data Based on Multiple Link Correlation in Feature Space", Proc. of 13th World Congress of ITS, London, CD-ROM（Oct. 2006）
- (7) 森田哲郎、矢野純史、香川浩司、「エージェント協調を用いたプロブカー・データによる空間補完システム」、信学技報、ITS2009-33, pp.91-96（2009-12）
- (8) T. Morita, J. Yano, K. Kagawa, "Interpolation System of Traffic Condition by Estimation/Learning Agents", in Proceedings of 12th international Conference on Practice in Multi-Agent Systems, Nagoya, pp. 487-499（Dec. 2009）
- (9) T. Morita, J. Yano, K. Kagawa, "Multiagent Based Interpolation System for Traffic Condition by Estimation/Learning", Proc. of the 9th International Conference on Autonomous Agents and Multiagent Systems, Toronto, pp. 1697-1704（May 2010）
- (10) 足立堅一、多変量解析入門、篠原出版社、東京（2005）

執筆者

森田 哲郎：情報通信研究所 ICT 応用研究部  
主席 工学博士  
交通情報補完システムの開発に従事

