

## 決定木を利用した交通渋滞予測手法に関する考察

金澤 明浩<sup>†</sup> 杵渕 哲也<sup>†</sup> 毛利 仁士<sup>‡</sup> 小川 智章<sup>‡</sup> 市河 研一<sup>†</sup> 荒川 賢一<sup>†</sup>

<sup>†</sup> 日本電信電話株式会社 NTT サイバースペース研究所 神奈川県横須賀市光の丘 1-1

<sup>‡</sup> 東日本電信電話株式会社研究開発センター 東京都武蔵野市緑町 3-9-1

<sup>‡</sup> 東日本電信電話株式会社神奈川支店 神奈川県横浜市西区みなとみらい 4-7-3

E-mail: <sup>†</sup> {kanazawa.akihiro, kinebuchi.tetsuya, ichikawa.kennichi, arakawa.kenichi}@lab.ntt.co.jp

<sup>‡</sup> h.mouri@rdc.east.ntt.co.jp <sup>‡</sup> tomoaki.ogawa@east.ntt.co.jp

あらまし 決定木を用いた交通渋滞予測手法を提案する。VICS(Vehicle Information and Communication System)交通情報を対象とした予測精度評価実験を行い、従来手法である現在の渋滞状況をそのまま将来の予測渋滞状況とする持続予測手法、および曜日時間帯毎の最頻渋滞状況を予測渋滞状況とする統計予測手法との比較を行った。その結果、本手法は、商用利用可能な計算コスト内で、予測先時間によらず従来手法より高精度に将来の渋滞状況を予測する事が可能であることを示した。

キーワード 決定木, VICS 交通情報, 渋滞予測

## Preliminary Consideration of using Decision Tree to Predict the Future Traffic Congestion

Akihiro KANAZAWA<sup>†</sup> Tetsuya KINEBUCHI<sup>†</sup> Hitoshi MOURI<sup>‡</sup> Tomoaki OGAWA<sup>‡</sup>

Kennichi ICHIKAWA<sup>†</sup> and Kenichi ARAKAWA<sup>†</sup>

<sup>†</sup> NIPPON TELEGRAPH AND TELEPHONE CORPORATION NTT Cyber Space Laboratories

1-1, Hikarinooka, Yokosuka-Shi, Kanagawa, 239-0847, Japan

<sup>‡</sup> NIPPON TELEGRAPH AND TELEPHONE EAST CORPORATION Research and Development Center

3-9-1, Midorichou, Musashino-Shi, Tokyo, 180-0012, Japan

<sup>‡</sup> NIPPON TELEGRAPH AND TELEPHONE EAST CORPORATION Kanagawa Branch

4-7-3, Minatomirai, Nishi-Ku, Yokohama-Shi, Kanagawa, 220-0012, Japan

**Abstract** This paper proposes a traffic congestion prediction method that is based on the use of decision trees. An experiment conducted using information from the VICS (Vehicle Information and Communication System) shows that the method is both accurate and fast; the method supports a wide range of look-ahead periods. A comparison against the continuous prediction method and the statistical prediction method confirms the superiority of the proposed method.

**Keyword** Decision Tree, VICS Traffic Information, Traffic Congestion Prediction

### 1. はじめに

ITS をはじめとする交通情報産業の発展を担う重要技術として、将来の交通渋滞を予測する技術が注目を集めている。ドライバーは予測された交通渋滞情報を受け取ることで、渋滞を回避して運転することができるし、旅行業者は、無駄のない観光コースを作成することができる。交通管理者は、予測データをもとに交通量のコントロールを行い、渋滞の解消や事故防止などを図ることも可能であり、幅広い分野への応用が期待されている。筆者らは、以上のような背景のもと、出発時にN分先の予測渋滞地図をカーナビ装置やパソコンから確認するという利用状況を想定し(図1)、

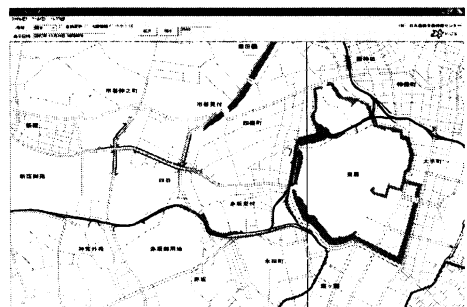


図1. 予測渋滞地図<sup>\*1</sup>

ATIS(Advanced Traffic Information System)交通情報[1][2]を利用した渋滞予測技術の研究開発を進めてきており、これまでに交通渋滞予測手法の一つとして事例ベース推論方式による渋滞予測方式[3]を提案してきた。

注\*1) 図1の予測渋滞地図における渋滞情報は、JARTIC(財)日本道路交通センター、VICS(財)道路交通情報システムセンター提供の渋滞情報を利用して作成した。背景の地図は、(財)日本デジタル道路地図協会のDRMデータを使用した。

## 2. 従来手法の問題点

これまで提案されている従来の交通渋滞予測手法は、大別して、道路モデルに基づく交通流シミュレーションによる予測手法と、過去における道路の渋滞変化パターンの時間変化を観測し、これを統計的に処理して将来の渋滞状況変化を予測する統計的手法に大別される。全国規模の一般道路における渋滞予測への適用を考慮すると、流入ポイントにおける車両流入量の特定が困難な一般道路では、前述のシミュレーションによる予測手法は現実的ではない。

そこで高速道路・一般道路の道路種別に関わらず適用可能な渋滞予測手法を考えると、統計的手法をベースとした以下のような予測方式が挙げられる。

### (1) 持続予測方式

現在の渋滞状況が短時間先の将来においても持続するという仮説の基に、将来の渋滞状況として予測処理時点の渋滞状況を提示する方式。0~15分程度の短時間先の渋滞予測においては予測精度が高く、突発的な事故の発生や大雨等の異常気象による渋滞状況の変動への適応性も高く、予測値(予測渋滞パターン)を算出するための計算機の処理コストが小さくて済むという長所がある。反面、予測先時間が長くなるにつれて、予測精度が著しく低下するという短所がある。

### (2) 曜日・時間帯別統計最頻パターン方式

過去における各道路の渋滞状況パターンを曜日・時間帯別に分類・集計し、その中における最頻渋滞パターンを予測先時刻における予測渋滞パターンとして提示する方式。予測処理時における道路状況に関する情報を必要としないため、数日、数週間、数ヶ月先の長期予測が可能であり、予測先時間によらず予測精度が一定である。また予測値算出の計算機コストも比較的小さくて済むという特長があるが、予測処理時の道路状況を考慮していないがために、突発的な事故の発生等による渋滞状況の変動への適用ができない(事故による渋滞が発生しても予測先時刻における通常時の渋滞状況が「渋滞なし」であれば、渋滞しないという予測結果が提示されてしまう)という欠点がある。(以降、

本手法が渋滞状況パターンの統計最頻値を利用しているという意味で、狭義の「統計予測方式」、「統計予測」と呼ぶこととする)

### (3) 事例ベース推論方式

事例ベース推論による具体的な渋滞予測手法については文献[3]に譲ることとして、ここではその特徴を以下に述べる。事例ベース推論方式は、予測精度の面では前述の持続予測方式、統計予測(曜日・時間帯別統計最頻パターン)方式と比較して、短時間先から長時間先まで安定してよい予測精度を保つとともに、突発的要因(事故、異常気象)による渋滞変化への適応も出来、また予測因子としての特徴量を、曜日、時刻のみとすることで前述の統計予測と同じ予測結果を得ることが可能であり、数日、数週間先の長期予測にも適用できるという利点を持つ。しかし、事例ベース推論による渋滞予測方式においては、膨大な過去の渋滞事例パターンと予測処理時の現在の渋滞状況パターンとの、特徴量空間における多次元ベクトル間のユークリッドノルムの算出処理、および最も距離が小さい渋滞パターン事例の比較選別処理を、渋滞予測処理時に動的に実行する必要があるため、日本全国を対象とした全道路の渋滞予測情報をリアルタイムに得ようとする、計算機の処理コストが極めて大きくなるという欠点がある。

そこで我々は、渋滞予測技術の研究開発の次ステップとして、日本全国の高速道路・一般道路を対象とした実利用可能な渋滞予測技術の開発を念頭に、取り扱う道路交通情報として、日本全国のリアルタイム交通情報をカバーし、センターから直接あるいは情報提供事業者を経由して入手可能なVICS(Vehicle Information and Communication System)交通情報[4][5]をターゲットに変更するとともに、事例ベース推論方式と同等あるいはそれ以上の予測精度が得られ、かつ実時間で商用利用可能な計算コストで予測情報を得られるような交通渋滞予測方式の確立を目標とした。

## 3. 決定木による渋滞予測

### 3.1. VICS 交通データの特徴

本論文でターゲットとしたVICS交通情報として提供される各道路の渋滞状況や道路リンクの通過旅行時間情報は、高速道路については各高速道路の道路管理者が、また一般道路については主に各都道府県の警察が設置したセンサから得られる情報であり、従来我々が研究対象としてきたATIS交通情報と同じである。その意味でVICS交通情報に含まれる渋滞情報の特徴は文献[3]に示した通りであるが、確認のために改めてVICS交通情報を対象とした調査結果を以下に再掲する。調査対象は、2001.2.1~2002.3.31の期間に提供さ

れた神奈川県内の VICS 道路リンク約 4,300 を対象とした。これは、提供されるべき交通情報標本数 122,112(2001.2.1~2002.3.31 の 5 分間隔で得られる交通情報事例数)の内、約 98%に当たる 12,000 事例数以上の実データが提供されている道路リンクである。

(1) 3段階に離散化された渋滞度

VICS 交通情報として提供される渋滞情報は図3に示すように、各道路毎に渋滞状態を定義する複数の道路区間(以降、リンク素片と呼ぶ)毎に、その渋滞度を(1=渋滞なし、2=混雑、3=渋滞)という3段階に離散化された値で提供されている。

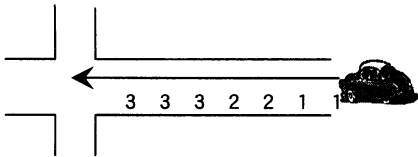


図2 提供される各リンク素片の渋滞度

(2)渋滞事例数の少なさ

提供された交通情報事例数 122,112 に対する混雑・渋滞情報が含まれている事例数の割合を調べた。図3にヒストグラムとして結果を示す。これから、大部分の道路リンクにおいて、混雑・渋滞している事例数が全交通情報提供事例数の約 5%以下であり、渋滞事例が稀な事象であることがわかる。

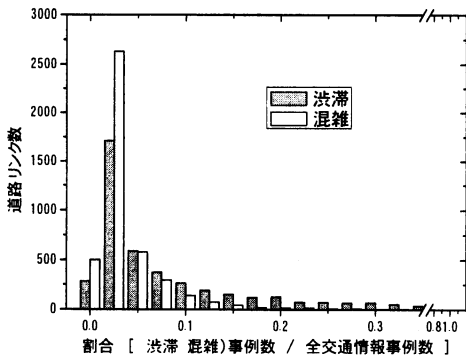


図3 全事例数に占める渋滞混雑事例数の割合

(3)渋滞状況パターンの実現値の少なさ

渋滞度は上記(1)で述べたように3段階に離散化されているため、n個の道路リンク素片からなる道路リンクがあるとすると、リンクの渋滞状況は、理論的には3^n種類のパターンがある。しかし実際のデータでは、概ね長さ3以上のリンクに対して、2~10数種類のパターンしか観測されていない。これは、平均して理論値の約 8.7%程度であり、現実には、非常に少ない数

の渋滞パターンしか発生していないことがわかる。

3.2. 新たな渋滞予測方式の検討

ここで新たな渋滞予測方式の検討に当たり、従来の予測手法の内、予測精度に関する性能が最も良かった事例ベース推論方式について再考してみると以下のようことが言える。

事例ベース推論方式[3]においては、まず時系列データ中のある時刻における各道路リンクの渋滞状況パターンを、当該道路の各リンク素片の渋滞度を示す数値からなる多次元ベクトルの形で定量化する。そして予測処理における予測因子としての「曜日」とか「時間帯」といった非計量的な因子については、当該曜日・時間帯における渋滞状況パターンの多次元ベクトルの平均値をその特徴量としてこれを定量化する。こうして渋滞状況パターンの予測問題を、離散値(一部実数値)の特徴ベクトルからなる時系列データの変化パターンの認識問題として置き換え、認識処理で用いられる k-NN(Nearest Neighbor)法と同様の原理を適用することで、将来の渋滞状況パターンを予測している。

ところで、上記のような「曜日」等の非計量的データを対象としたパターン認識の手法として「決定木(Decision Tree)」[6]が知られている。決定木によるパターン認識処理の特徴としては次のような事項が挙げられる。

- (a) 実数値や離散値等の計量可能なデータばかりでなく色(赤,青,黄色)、曜日(日,月,...土)といった、非計量的データを統一的に扱うことが可能。
- (b) パターン認識のための認識辞書の作成(いわゆる学習処理)のための計算コストは大きい。認識対象標本データに対するパターン認識処理の計算コストは非常に小さい(木構造データベースの検索処理コストと同等)。

我々は上記のような決定木の特徴に着目し、これを渋滞状況パターンの予測処理に適用する方式を検討した。決定木による渋滞予測処理イメージを図4に示す。

現在のデータ  
曜日:月 時刻:7:30 現在の渋滞状況:渋滞なし

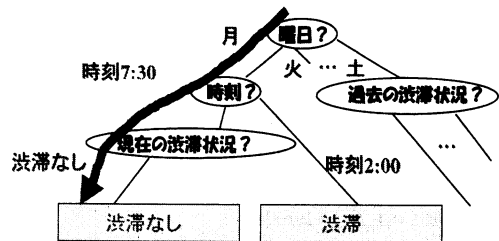


図4 決定木による渋滞予測処理イメージ

### 3.3. 決定木を利用した渋滞予測手法

本節では、決定木によるパターン認識手法を渋滞予測に適用する際の具体的な処理手順を順に述べる。

#### 3.3.1. 予測対象データ

VICS 交通情報から提供される渋滞情報は、3.1.(1)に示した通り、道路リンク内の道路リンク素片単位にその渋滞度を表す数値(1=渋滞なし, 2=混雑, 3=渋滞)の組の時系列データとして与えられる。問題は、このような「渋滞度を表す数値の組」で表された過去の時系列データを学習データとして、将来における「渋滞度を表す数値の組」を予測することとなる。しかし道路の渋滞状況をこのような「渋滞度を表す数値の組」としてそのまま扱うと、処理系が複雑となることから、我々は、3.1.(3)にあるように実際に出現する渋滞状況パターンの出現値が少ないことに着目し、学習データに含まれる過去の渋滞パターンの全てに「渋滞状況パターン番号」を付与し、この「渋滞状況パターン番号」の時系列データを予測処理の対象データとした。図5に渋滞状況パターン番号の時系列データのイメージを示す。

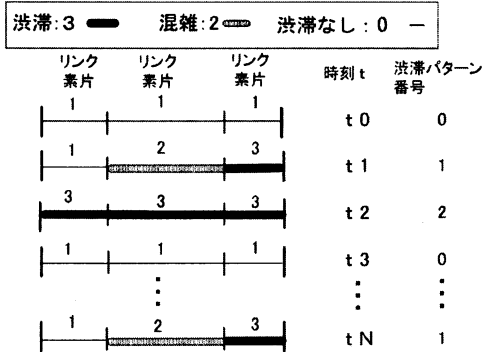


図5 渋滞パターン番号付与例

#### 3.3.2. 決定木の分岐を決める特徴量

前節より得られた予測先時刻の渋滞状況パターン番号の集合を、ある特徴量によって逐次分類し、予測のための決定木を作成することになる。ここで重要となるのは、評価する特徴量として何を選定するかである。本論文では、事例ベース推論による渋滞予測と予測精度を比較するという意味で、今回は以下に示す5つの項目を特徴量として採用した。

- ①曜日：0=月, 1=火, …, 6=日 & 休日,
- ②時刻：0, 1, 2, …, 286, 287  
(一日の時刻を5分単位に番号付け)
- ③五十日：0=五十日でない, 1=五十日
- ④現在の渋滞状況パターン番号：0, 1, 2, …

⑤現在より1~Nstep前の渋滞状況パターン番号  
学習データとして過去一年分の渋滞事例を学習データとすると、それは図6に示す通りとなる。

なお、特徴量⑤として過去何ステップ分の渋滞状況パターンを評価すればよいか。また学習データの期間としてはどの程度(1, 3, 6, 12ヶ月)用意すればよいか、についてはこれらを幾つか変更して作成した決定木による数値実験により検討した(第4章)。

0	0	0	0	1
1	0	0	0	2
2	0	0	1	1
...	...	...	...	...
287	0	0	0	0
0	1	1	0	0
1	1	1	2	1
2	1	1	1	1
...	...	...	...	...
286	1	1	1	1
287	1	1	0	0
0	0	2	0	0
1	0	2	2	1
2	0	2	1	1
...	...	...	...	...
...	...	...	...	...

288×365日  
約1万事例数の  
学習データ

図6 学習データ例

#### 3.3.3. 決定木の作成方法

前節で作成した学習データから、情報量に着目して決定木を作成する。決定木作成方法には CART、ID3、C4.5 など様々なもの[6][7][8]がある。いずれの決定木作成方法においても重要な点は以下の2つである。

- (a) 木分割の評価基準
- (b) 木分割の停止・枝刈り

##### 3.3.3.1. 木分割の評価基準

木の分割を評価する基準としては、情報量利得、情報量利得比、ジニ不純度等[6]が用いられるが、本論文では、決定木の分岐を決める特徴量として採用した①曜日, ②時刻, ③五十日, ④現在の渋滞状況パターン番号, ⑤過去の渋滞状況パターン番号, のそれぞれがとるクラス数にばらつきがあることから、文献[9]を参考にし、情報量利得比を分割のための基準とした。

以下に情報量利得比を基準とする決定木の分割方法を簡単に説明する。

一般的に集合 S の情報量は、式(1)であらわされる。

$$info(S) = - \sum_{j=1}^k \frac{freq(C_j, S)}{|S|} \times \log_2 \left( \frac{freq(C_j, S)}{|S|} \right) \dots (1)$$

ここで、 $freq(C_j, S)$ は、集合 S におけるクラス  $C_j$  に所

属する事例数であり、 $|S|$ は集合  $S$  の全事例数の事である。

一般的にある集合  $S$  を、ある条件  $X$  で複数の集合  $S_i$  に分割した後の情報量は式(2)であらわされる[9]。

$$\text{info}_X(S) = - \sum_{j=1}^n \frac{|S_j|}{|S|} \times \text{info}(S_j) \dots (2)$$

次に、集合  $S$  の情報量  $\text{info}(S)$  の定義から、 $S$  を  $n$  個の部分集合へ分割すること自体によって得られる情報量(分割情報量)を次式により定義する[9]。

$$\text{split\_info}(X) = - \sum_{j=1}^n \frac{|S_j|}{|S|} \times \log_2 \frac{|S_j|}{|S|} \dots (3)$$

分割前の情報量と分割後の情報量との差(情報量利得)  $\text{Gain}(X)$  は、(2)式より、次式の通り与えられる。

$$\text{Gain}(X) = \text{info}(S) - \text{info}_X(S) \dots (4)$$

よって、情報量利得の分割情報量に対する比(情報量利得比)  $\text{Gain ratio}(X)$  は、

$$\text{Gain ratio}(X) = \text{Gain}(X) / \text{split\_info}(X) \dots (5)$$

となる。

今、図6に示した学習データのうち、数字が明確に記載されている12事例のみから、学習データが構成されているものと仮定する。この12事例の学習データを集合  $S$  とする。この時、予測先時刻における渋滞パターン番号は、0が5事例、1が6事例、2が1事例であるので、この分割されていない集合  $S$  の持つ情報量  $\text{info}(S)$  は、

$$-\frac{4}{12} \log \frac{4}{12} - \frac{7}{12} \log \frac{7}{12} - \frac{1}{12} \log \frac{1}{12} = 1.28$$

である。

これを時刻  $t$  という特徴量で  $S_{10}, S_{11}, S_{12}, S_{1286}, S_{1287}$  に分割した時、部分集合  $S_{10}$  の情報量  $\text{info}(S_{10})$  は、

$$-\frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} = 0.92$$

であり、それぞれ

$\text{info}(S_{11})=0.92$ ,  $\text{info}(S_{12})=0$ ,  $\text{info}(S_{1286})=0$ ,  $\text{info}(S_{1287})=0$  となる。

従って、分割後の情報量  $\text{info}_t(S)$  は式(2)より

$$\frac{3}{12} \times 0.92 + \frac{3}{12} \times 0.92 + \frac{3}{12} \times 0 + \frac{1}{12} \times 0 + \frac{1}{12} \times 0 = 0.46$$

となる。又、分割情報量  $\text{split\_info}(t)$  は、式(3)より、

$$\frac{3}{12} \log \frac{3}{12} + \frac{3}{12} \log \frac{3}{12} + \frac{3}{12} \log \frac{3}{12} + \frac{1}{12} \log \frac{1}{12} + \frac{1}{12} \log \frac{1}{12} = 2.10$$

となる。

よって、時刻  $t$  という特徴量を利用した分割による情報量利得比は、

$$\begin{aligned} \text{Gain ratio}(t) &= \frac{(\text{info}(S) - \text{info}_t(S))}{\text{split\_info}(t)} \\ &= (1.28 - 0.46) / 2.1 = 0.39 \end{aligned}$$

となる。同様に、各特徴量における情報量利得比は五十日による情報量利得比=0.04/0.99=0.39

曜日による情報量利得比=0.27/1.5=0.18

現在の渋滞パターン番号による情報量利得比=0.68/1.5=0.47

となるので、ここでは、現在の渋滞パターン番号による情報量利得比が最も大きくなり、この特徴量により木が分割される。

木の分割方法を情報量利得を基準とした場合と、情報量利得比を基準とした場合による予測精度の違いを4.3.3.に示す。

### 3.3.3.2. 木分割の停止・枝刈り

決定木の分割停止および枝刈りに関しては、文献[9]を参考に、先ず前述の分割方法を再帰的に行い、分割後の予測先時刻の渋滞パターンの集合が、全て同一の渋滞パターンとなるか、これ以上分割できないところまで木を分割し、その後、構築された木の枝刈りを行った。木の分割を最後まで行わず、ある条件により停止した場合には、地平線効果とよばれる弊害がおき、逆に最後まで分割を行った場合、決定木が学習データに過剰に適合してしまうといった問題が挙げられている[6]。決定木が学習データに過剰に適合するということは、最悪、1万事例の学習データから一万パターンの分岐条件を持つ決定木が作成されることを意味し、このような分岐条件が現実的に有効であるとは考えにくい。これらの問題を解決するために、枝刈り処理があり、決定木が学習データに過剰に適合するといったことを防いでいる。この枝刈りの有無による予測精度結果の違いを4.3.3.に示す。

## 4. 評価値実験

本章では、前述の決定木による渋滞予測手法を用いて、実際に VICS 交通情報を使用した評価実験を行った結果を示す。

先ず、評価実験を行うに当たっての諸条件について述べ、次に渋滞の予測精度に関する評価、予測結果を算出するのに要する処理性能、および決定木作成におけるパラメータを変化させた場合の予測精度への影響について順次述べる。

### 4.1. 評価実験に際しての諸条件

今回評価実験に用いたデータは、2001.2.1~2002.3.31の14ヶ月間に観測されたVICS交通情報で、

対象道路リンクは神奈川県内の VICS 道路リンクの内、JIS-X0410 に定める 2 次メッシュ番号 533904, 533905, 533914, 533915, 533924, 533925, 533934, 533935 の 8 メッシュ（おおよそ南北には横須賀～厚木、東西には、横浜～鎌倉を含むエリア）に含まれる 2,290 リンクの渋滞情報である。

## 4.2. 予測精度評価

### 4.2.1. 評価基準

予測精度評価は、持続予測方式、統計予測方式、及び決定木方式のそれぞれについて、予測用データの学習期間を 2001.2.1～2002.1.31 の一年間とし、評価対象期間を 2002.2.1～2002.2.28 の一ヶ月間として、各予測方式による予測結果と当該期間中の実測渋滞状況をリンク素片毎に比較することで行うこととし、予測結果をその“あたり”、“はずれ”で分類した(図 7)。

◆リンク素片毎に予測結果を評価する

予測値	2	2	3	3	3	0	0
実測値	2	1	1	2	0	0	1



◆リンク素片毎に予測結果を分類

予測値 実測値	渋滞:3	混雑:2	渋滞なし:1
渋滞:3	A あたり	B はずれ	C はずれ
混雑:2	D はずれ	E あたり	F はずれ
渋滞なし:0	G はずれ	H あたり	I あたり

図 7 予測結果の分類

本論文では予測結果を評価するための指標として以下の 4 つを定義した。

(i) 体的中率  $Pa = (A + E + I) / (A + B + C + D + E + F + G + H + I)$

(ii) 的中率  $Pb = A / (A + D + G)$

(iii) 渋滞再現率  $Pc = A / (A + B + C)$

(iv) 渋滞なし的中率  $Pd = I / (G + H + I)$

ここで  $Pa$  は全体での予測的中率を示す。一般の予測にあたり/はずれはこの  $Pa$  が問題となるが、本論文における実データは 3.1.(2) で述べたように、全体の事例の内、渋滞なしの事例が約 95% である。このため  $Pa$  のみで評価すると常に渋滞なしを出力する予測に対して高い評価を与えてしまうことになる。

このような場合は  $Pb, Pc, Pd$  を考える。それぞれの意味は次の通りである。

・渋滞的中率：渋滞と予測した事例数に対する、実際に渋滞が発生した事例数の割合。“渋滞する”という予測がどれだけ当たったかを示す。

・渋滞再現率：実際に渋滞が発生した事例について、どの位その渋滞の発生を予測できたかという事例数の割合。

・渋滞なし的中率：渋滞しないと予測した事例数に対する、実際に渋滞しなかった事例数の割合。

なお、本論文では、リンク素片毎のあたり/はずれの結果に対して、その事例数をそのまま計算するのではなく、各リンク素片の長さで重み付けした値で予測精度を算出した。

### 4.2.2. 評価結果

前節の評価基準に従って、持続予測方式、統計予測方式および決定木予測方式の 3 つについて、予測先時間毎の体的中率、渋滞的中率、渋滞再現率、渋滞なし的中率を、それぞれ図 8, 図 9, 図 10, 図 11 に示す。

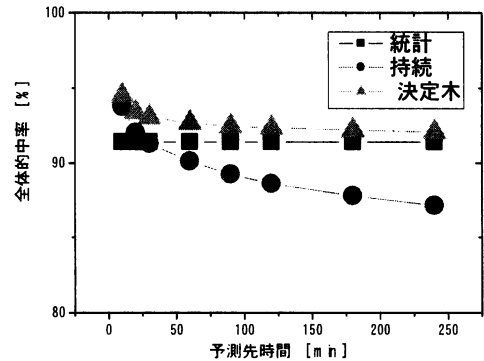


図 8 体的中率と予測先時間

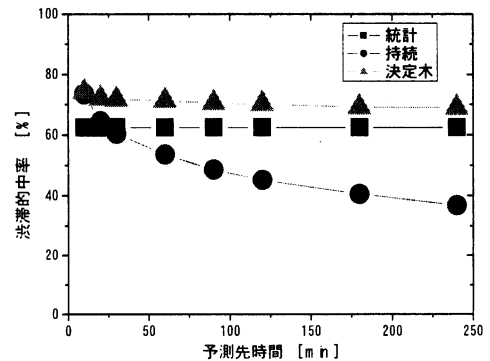


図 9 渋滞的中率と予測先時間

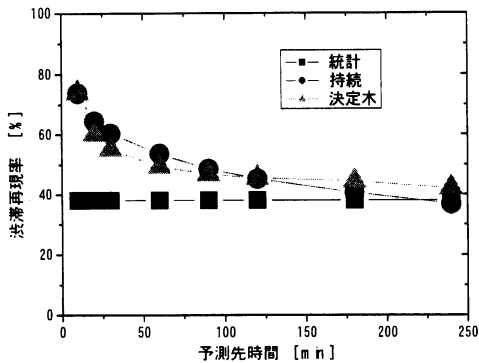


図 10 汎滞再現率と予測先時間

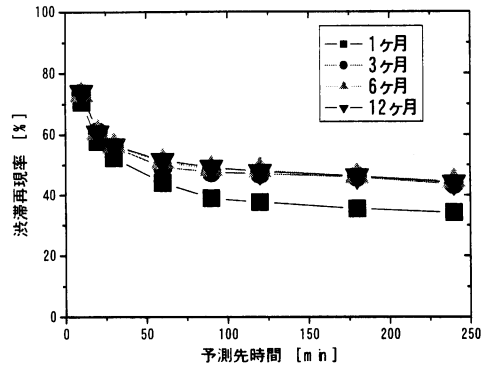


図 12 学習期間と汎滞再現率の関係

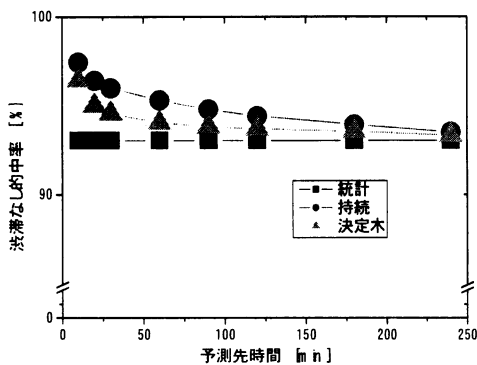


図 11 汎滞なしの中率と予測先時間

汎滞的中率  $P_b$ 、汎滞再現率  $P_c$  とともに、予測先時間にかかわらず決定木による予測手法が高精度に予測できていることがわかる。また、統計予測手法は、汎滞的中率と比較して、汎滞再現率が極端に低いことがわかる。このことから、統計予測は、あまり汎滞と予測せず、より確実に汎滞するときのみ汎滞と予測する傾向があることがわかる。

### 4.3. 特徴量と予測精度

本節では、決定木作成にあたって用いる特徴量（予測因子）、および決定木作成方法の違いと予測精度との関連を評価実験により検証した結果を述べる。

#### 4.3.1. 学習期間

リンク単位の決定木による予測手法を実現するために、どれくらいの学習期間が必要かを調べた。学習期間を以下のようにして比較した結果を図 12 に示す。

- I. 1 ヶ月 2002.1.1 0:00 ~ 2002.1.31 23:55
- II. 3 ヶ月 2002.11.1 0:00 ~ 2002.1.31 23:55
- III. 6 ヶ月 2002.8.1 0:00 ~ 2002.1.31 23:55
- IV. 12 ヶ月 2001.2.1 0:00 ~ 2002.1.31 23:55

この結果から、学習機関として最低 3 ヶ月あれば十分であることがわかる。なお汎滞的中率  $P_b$  も同様の傾向であった。

#### 4.3.2. 過去の汎滞パターン

現在から過去何ステップ前の汎滞状況までを考慮すべきかを検討するために、特徴量として採用する過去の汎滞状況ステップ数を変化させて、その予測精度に対する影響を調べてみた。以下に 30 分先の汎滞的中率  $P_b$  および、汎滞再現率  $P_c$  の値を示す。

- I. 現在の汎滞状況 ( $P_b, P_c$ ) = (69.3, 52.7)
- II. 現在 + 5 分前 ( $P_b, P_c$ ) = (71.6, 55.1)
- III. 現在 + 5, 10 分前 ( $P_b, P_c$ ) = (69.1, 53.2)
- IV. 現在 + 5, 10, 15, 20, 25 分前 ( $P_b, P_c$ ) = (69.6, 53.6)

特徴量として採用する  $N$  ステップ前の汎滞事例を増やしていても、予測精度が向上する傾向は特になく、現在の汎滞状況だけを考慮すれば充分であることがわかった。

#### 4.3.3. 決定木作成方法の違い

決定木は利得基準を用いて作成するか、利得比基準を用いるかで木の構成が異なる。利得比基準を用いた場合の方が優れていると一般的に考えられている[10]が、実際に、それぞれの作成方法で決定木を作成し、予測を行ったところ以下ようになった。30 分先の汎滞的中率  $P_b$  と汎滞再現率  $P_c$  は、

- I. 利得比基準 ( $P_b, P_c$ ) = (71.7, 55.1)
- II. 利得基準 ( $P_b, P_c$ ) = (71.2, 55.6)

となり、予測精度として大きな差はなかった。

また、決定木の枝刈りの有無による予測精度の違いについては、予測先時間が大きい時間帯で枝刈りをしなかった場合の方が、的中率  $P_c$  が低く再現率  $P_c$  が高かった。30 分先と 240 分先の  $P_b, P_c$  を示す。

- I. 枝刈り有り
  - 30 分先 ( $P_b, P_c$ ) = (71.7, 55.1)
  - 240 分先 ( $P_b, P_c$ ) = (68.9, 42.0)

## II 枝刈り無し

30分先 (Pb,Pc)=(68.2,56.0)

240分先 (Pb,Pc)=(64.2,45.8)

### 4.4. 予測処理性能

決定木方式による渋滞予測性能について、先に 3.1 で取り上げた神奈川県内の約 4,300 リンクについて、予測先時間として、それぞれ 10,20,30,60,90,120,180,240 分先の渋滞状況を予測する決定木を作成し、その予測処理時間を測定した。測定結果を表 1 に示す。なお予測処理性能の測定に用いた計算機 (1 台) のスペックは以下の通りである。

- ・ CPU : Xeon 2.8GHz × 2
- ・ メモリ : 4Gbytes
- ・ DISC : Ultra Wide SCSI 160MHz 140Gbytes × 6
- ・ OS : RedHat Linux 7.3

表 1 予測先時間数と決定木のサイズ/予測処理時間

予測先時間数	決定木サイズ (MB)	処理時間 (秒)	予測先分の組み合わせ
1	2,744	—	10
2	4,646	105	10,20
4	7,134	162	10,20,30,60
6	8,373	213	10,20,30,60,90,120
8	9,395	290	10,20,30,60,90,120,180,240

### 5. まとめと今後の課題

VICS 交通情報を用いて、決定木を用いた交通渋滞予測手法に関する評価実験を行い、本手法を用いることで広範囲・高精度・高速な予測渋滞箇所地図を作成することが可能であることを示した。神奈川県的一般道路を対象とした評価実験結果を以下にまとめる。

- (1) 決定木を作成し予測を実施するのに必要な学習期間は、最低約 3 ヶ月であることを確認した。
- (2) 現在の渋滞状況のみを特徴量として考慮し、5 分、10 分前の渋滞状況は考慮せずとも、十分な予測精度が得られることを確認した。
- (3) 利得比基準と利得基準の 2 通りで決定木を作成し、数値実験をおこなったところ、予測精度に関して特に差がないことがわかった。また、決定木作成時に、枝刈りを行わなかった時と、枝刈りを行った場合で予測精度を比較したところ、予測先時間が長い時には、枝刈りを行わなかった時の方が再現率が高かった。
- (4) リンク単位の決定木による渋滞予測手法が、予測時間の見地から、商用可能な計算コスト内である

ことが確かめられた。

今後の課題としては、的中率および再現率が低い道路リンクを対象に、交通渋滞特性や作成した決定木の構造を解析し、決定木作成アルゴリズムにも改良を加えていく予定である。また予測するための特徴要因として考慮していない上流リンクの渋滞状況などを考えていく予定である。

また、実際に商用利用を開始するときに重要となってくる近隣の道路状況が極端に変わった場合への対処方法、新たに提供を開始された学習期間が全くない道路リンクの予測方法なども、今後検討していきたいと考えている。

### 文 献

- [1] 堀内雅人, 伊藤太郎, 丹後 健, 木野村雅, 大内敏弘, 交通情報提供サービス(ATIS), "電学道路交通研資," RTA-95-12, pp.97-106, 1995.
- [2] <http://www.atis.co.jp/>
- [3] 毛利仁士, 堀越 力, 小川智章, 安達文夫, 鈴木智, "事例ベースによる交通渋滞予測手法," 電子情報通信学会論文誌 B, Vol.J82-B, No.11, pp.1993-2001, 1999.
- [4] <http://www.vics.or.jp/>
- [5] 田中正晴, "カーナビが双方向情報端末に進化," 日経エレクトロニクス, no.662, pp.91-106 1996.
- [6] 尾上守夫, "パターン識別," 新技術コミュニケーションズ, 2001..
- [7] Ruggieris, S. "Efficient C4.5," IEEE Trans Knowl Data Eng, Vol.14, pp.438-444, 2002.
- [8] Mcsherry, D. "Strategic induction of decision trees," Knowl Based Syst, Vol12, pp.
- [9] Quinlan, J.R., "C4.5 Programs For Machine Learning," 1993
- [10] Quinlan, J.R., "Decision trees and multi-valued attributes," Machine Intelligence, 11(pp.305-318), 1988.