

プローブカーデータによる時空間道路交通データベースの構築および自動更新手法  
Updating Methodology of Spatio-Temporal Database of Probe-Vehicle Travel Time.

薄井 智貴† 南 宏樹†† 三輪 富生††† 森川 高行†††† 山本俊行†††††  
Tomotaka Usui Hiroki Minami Tomio Miwa Takayuki Morikawa Toshiyuki Yamamoto

名古屋大学大学院 環境学研究科

## 1. はじめに

急速なIT技術の進展とともに高度道路交通システム(ITS)の実用化が進められ、道路交通情報の収集、生成も進化が遂げられつつある。従来、光ビーコンやトラフィックカウンター等を用いたVICS情報が主流であった。しかし、VICSは定点観測であるため、路側観測機を有する道路において確実な情報を得られる反面、細街路や郊外路線など非設置区域においては、何ら情報を得ることができず、またそれらに対応するためには膨大なコストがかかる。そこで、都心細街路での交通情報をリアルタイムに収集可能な、プローブカーを用いた情報システムの開発が注目を集めている。

このプローブカーシステムは自動車一台一台にセンサを取り付け、車両の位置情報や車両挙動などのモニタリングするものであり、得られた情報を加工、蓄積することにより、道路混雑情報、旅行時間、降雨情報などをリアルタイムに収集、提供することを可能にする。また、路側観測機などのハード的インフラを必要としないため比較的安価に、エリアを限定することなく情報収集が可能である。しかし、収集されるデータは時間的、空間的密度が低く、データ量や情報のばらつきも大きい。より精度の高い交通情報を提供するためには、ある程度長期間のデータ収集、蓄積が必要となる。さらに、日々刻々と変化する交通状況に対し、現在の状況を反映していない過去の古い情報すべてを長期間蓄積することは、必ずしも適切な情報の生成とならず、また膨大なデータソースの損失にもつながる。そのため、新しく収集したプローブカーデータを必要な情報に加工、蓄積し、より精度の高い時空間データベースを作成する必要がある。

そこで本研究では、プローブカーデータを活用した現況の旅行時間情報提供のための時空間データベースを構築し、常に最新の情報を提供するための自動更新手法について4つの方法で検討を行った。さらに、構築したデータベースと実測値データを比較することにより、更新手法の優位性を示す。

## 2. プローブカーデータによる時空間データベース

プローブカーデータは、個々の車両の走行データであり、日々の膨大なデータを処理する必要がある。そのため、得られたデータを効率よく、かつ効果的に管理するため

のデータベース構築が不可欠である。本研究では、時間的、空間的なプローブカーデータを取り扱うための時空間道路交通データベースを構築し、収集したプローブデータから得られるリンク旅行時間を時間帯別にデータベースに格納した。

## 3. データベースの構築と自動更新手法

構築したデータベースにプローブカー情報を蓄積し、常に精度の高い旅行時間を提供するための4つのデータベース更新手法を比較する。1つ目は過去に蓄積したプローブカーデータをすべて利用し、リンク毎にある一定の時間間隔で平均旅行時間を算出しデータベース化したものである (Pattern A)。この手法は、長期データを一括処理しデータベースを構築、更新するため、膨大な蓄積データを処理するため処理コストも高い。次に、取得したプローブカーデータのうち、ある地点から最近までのデータを利用する手法で、2つのパターンを想定した。1つは、リンク毎に最近のプローブカーデータの一定数 (50, 100, 200 data) を平均旅行時間としたもの。もう1つは、同じくリンク毎に最近数ヶ月分 (2, 4ヶ月) のプローブカーデータを集計したものである (Pattern B, C)。この結果により、プローブカーデータ数、あるいは収集期間によって旅行時間精度がどの程度誤差があるか確認できる。以上に示した3つの手法はいずれもリンク旅行時間の分布の単峰性を仮定している。しかし、道路リンクを走行する信号の有無によりその分布が多峰性となる場合が多い。そこで5つ目のパターンとして、データ収集と平行してデータが従う分布形を逐次的に学習し、学習した確率密度分布によりリンク旅行時間を確定する忘却型学習アルゴリズムを用いる (Pattern D)。本稿では、紙面の都合上、Pattern Dの手法について詳しく述べることにする。

## 4. 忘却型学習アルゴリズム

日々刻々データが収集されるプローブデータをデータベース化する場合、過去にデータ収集を開始した時点から現在まですべてのデータを蓄積するには膨大な記憶領域が必要となりデータソースの無駄となる。そのため、新規観測データを反映させると同時に、現況交通情報生成に影響のない過去のデータを削除することは必要不可欠である。また、先に述べたように、データベースを新規プローブカーデータにより更新する場合、多峰性を考慮して旅行時間を集計する必要がある。そこで本研究では、ある確率密度関数を学習し、学習したリンク旅行時間によりデータベースを更新する手法を提案する。本手法では、確率変数  $x$  が従う確率密度関数として、式(1)(2)で表される混合ガウス分布を用いる。

† 情 (修) 名古屋大学大学院 環境学研究科

†† 工 (学) 名古屋大学大学院 環境学研究科

††† 博 (工) 助教 名古屋大学大学院 工学研究科

†††† Ph.D. 教授 名古屋大学大学院 環境学研究科

††††† 博 (工) 准教 名古屋大学大学院 工学研究科

$$p(x|\theta) = \sum_{i=1}^k c_i p(x|\mu_i, \sigma_i^2) \quad (1)$$

$$p(x|\mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right\} \quad (2)$$

ここに、 $k$ は混合する分布の数で、 $c_i > 0$ 、 $\sum c_i = 1$ を仮定する。このとき、ある最新の(時点 $t$ の)旅行時間 $x_t$ が得られたとき、以下のSDEMアルゴリズム<sup>1)</sup>により、データが従う確率密度関数を学習する。

《SDEMアルゴリズム》 ( $r, \alpha, k$ : given)

Set  $\mu_i^{(0)}, c_i^{(0)}, \bar{\mu}_i^{(0)}, \Lambda_i^{(0)}, \bar{\Lambda}_i^{(0)} (i=1, \dots, k)$

Read  $x_t$

for  $i=1, \dots, k$

$$\gamma_i^{(t)} = (1-\alpha r) \frac{c_i^{(t-1)} p(x_t|\mu_i^{(t-1)}, \sigma_i^{2(t-1)})}{\sum_{i=1}^k c_i^{(t-1)} p(x_t|\mu_i^{(t-1)}, \sigma_i^{2(t-1)})} + \frac{\alpha r}{k} \quad (3.a)$$

$$c_i^{(t)} = (1-r)c_i^{(t-1)} + r\gamma_i^{(t)} \quad (3.b)$$

$$\bar{\mu}_i^{(t)} = (1-r)\bar{\mu}_i^{(t-1)} + r\gamma_i^{(t)} \cdot x_t \quad (3.c)$$

$$\mu_i^{(t)} = \bar{\mu}_i^{(t)} / c_i^{(t)} \quad (3.d)$$

$$\bar{\sigma}_i^{2(t)} = (1-r)\bar{\sigma}_i^{2(t-1)} + r\gamma_i^{(t)} \cdot x_t^2 \quad (3.e)$$

$$\sigma_i^{2(t)} = \bar{\sigma}_i^{2(t)} / c_i^{(t)} - \mu_i^{(t)2} \quad (3.f)$$

ここで、 $r (> 0)$ は忘却パラメータであり、値を大きくする程、過去のデータの影響を早く忘却する。パラメータ $\alpha$ は混合された分布の重み $c_i$ の推定値を安定化させる働きを持つ(本研究では1.0に設定)。 $c_i$ の初期値( $c_i^{(0)}$ )は、 $1/k$ とおく。

## 5. プローブデータへの適用

### (1) 分析対象ルートと比較データの概要

プローブカーデータは名古屋市周辺のタクシープローブデータで、観測された時空間情報をマップマッチングにより経路を特定し、旅行時間を推定する<sup>2)</sup>。データの収集期間は2002年1月~3月、10月~2003年3月、10月~2004年6月までの18ヶ月間である。比較対象ルートは、起終点を名古屋市西区の産業技術記念館と千種区の平和公園に設定し、平休日3日間のうちの上り下り全28ルートの実測走行結果と構築した旅行時間データベースから推定した予測結果を比較することにより、データベースの精度を検証した。

### (2) 分析結果

すべてのPatternについて、実測値との絶対誤差の平均を比較したところ、わずかではあるがPattern C(4ヶ月分)の誤差が一番少なく、つづいてPattern Aが少ないという結果となった。

さらに、各Patternとの比較では誤差に有意性がなかったが、忘却型学習アルゴリズムの有意性を示すため、一般のカーナビのデフォルト値(一般道路の旅行速度30km/h)による旅行時間予測値と忘却型学習アルゴリズムについて実測値との絶対誤差の差の検定を行ったところ、 $t$ 値が

4.31で両側有意確率0.0002となり、忘却型学習アルゴリズムが統計的に有意に誤差が小さいことが示せた。

また本手法の学習傾向をはかるため、あるリンクのリンク旅行時間分布を確認したところ、データ量の少ないリンクでは学習が十分に行われず、旅行時間の誤差が大きいことがわかった(図-1(a)(b))。

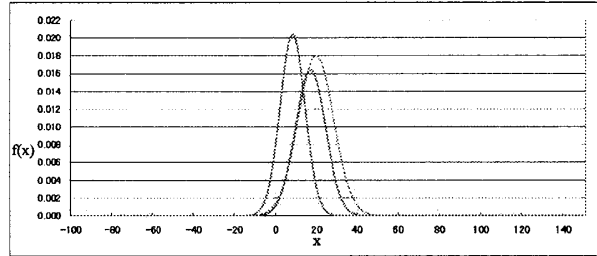


図-1(a) 学習が不十分なリンク旅行時間分布(20 data)

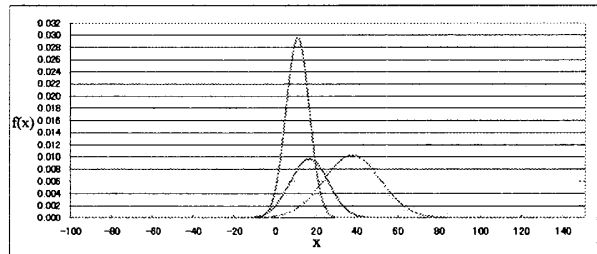


図-1(b) 学習が十分なリンク旅行時間分布(150 data)

## 6. まとめ

プローブカーデータを信頼性の高い交通情報生成に活用するための、データベース構築と更新手法を提案し、構築したデータベースによる推定旅行時間と実測旅行時間を比較した。その結果、すべてのデータを蓄積し旅行時間を算出したPattern Aが安定して実測値に近く、忘却型学習アルゴリズムより良い結果となった。ただし、忘却型学習アルゴリズムについては学習が不十分、つまりプローブカーデータ数が少ない場合は、初期パラメータの影響が大きくなり必ずしも正しい結果につながらないことも伺えた。

今後サンプル数を増やし、学習アルゴリズムの精度向上に向け、忘却パラメータや初期パラメータの最適な値を求めるとともに、データ収集頻度の低いリンクにおいても精度の高い旅行時間を更新できるよう最適なデータ数や集計する時間間隔を求めていきたい。

## 参考文献

- 1) Yamanishi, K. and Takeuchi, J.: A unifying framework for detecting outliers and change points from non-stationary time series data, Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2002.
- 2) 三輪富生, 木内大介, 山本俊行, 森川高行: “低コストプローブカーデータのオンラインマップマッチング手法”, 土木計画学研究・講演集, Vol.34, CD-ROM, 2006.