

プローブカーデータによる交通状況変化時点の検出に関する研究*

Detection of Traffic Situation Changing points Using Probe Vehicle Data *

三輪富生**・岡本和大***・山本俊行****・森川高行*****

By Tomio MIWA**・Kazuhiro OKAMOTO***・Toshiyuki YAMAMOTO****・Takayuki MORIKAWA*****

1. はじめに

プローブカーデータを活用した交通情報の生成が各地で盛んに行われ始めている。プローブカーシステムは大規模なインフラを必要としないため、空間的な自由度が高い。しかし、収集されるデータは時間的、空間的密度が低く、ばらつきも大きい。したがって、より精度の高い交通情報を生成するためには、ある程度長い期間にわたったデータ収集が必要となる。ここで、長期間にわたってデータを収集し、その平均速度や平均旅行時間等を用いる場合には、収集エリア内での一時的な路上工事等による渋滞や速度低下など、本来の交通状況を反映しないデータを除去する必要がある。さらには、車線幅等により交通状況が大きく変化した場合には、交通状況変化後のデータのみを用いて情報を生成する必要がある。

このように、現在の交通状況を表す情報を生成するためには、収集したプローブカーデータから交通状況が変化した時点を検出する必要がある。しかし、プローブカーデータには、車両間のばらつきやGPSエラーが含まれている。そこで、本研究では、このようなばらつきの大きなプローブカーデータから、交通状況の変化が発生した時点、すなわち交通状況変化時点を検出するための幾つかの手法を示す。さらに、シミュレーションデータや実際に収集されたプローブカーデータへの適用を通して手法間の比較を行う。これにより、交通状況変化時点の検出可能性を探ることを目的とする。

2. 変化時点検出手法

(1) 変化時点検出の考え方

プローブカーデータは個々の車両の走行データであり、その観測リンク旅行時間は車両間のばらつきを含む。例えば、同じ交通状況下であっても、信号交差点への僅かな流入時刻のずれから流出時刻が大きく変化することもある。さらには、GPSエラーやマップマッチングエラーは、このような情報のばらつきをさらに大きくするであ

ろう。したがって、分析に際しては、プローブカーデータの有するばらつきの考慮が必要となる。

図-1は、プローブリンク旅行時間と交通状況変化時点のイメージ図である。横軸をデータ収集日(時点)とし、300日目で変化時点が発生する例である。本研究では、同一曜日、同一時間帯において、長期間にわたって収集されたリンク旅行時間から、交通状況変化時点を検出することを考えるが、図に示すような突発的な交通状況変化を対象とする。すなわち、季節変動や人口増加等により徐々に変化する交通状況は考慮しない。

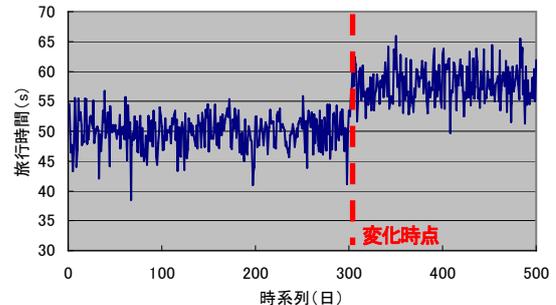


図-1 交通状況変化時点(イメージ)

(2) 変化時点検出手法

本研究では、5つの変化時点検出手法を示すが、これらは一定期間に収集されたデータに対し一括処理を行う2つの手法 (Batch Algorithm; Method I~II) と、データ収集と平行して逐次的処理を行う3つの手法 (Incremental Algorithm; Method III~V) に分けられる。前者は、多くのデータを扱うため情報量が多く、高い精度を期待できるが、処理量が膨大となり処理コストが高くなる。一方、後者は、新しく収集したデータによる情報の変化により検出を行うため、精度が低くなることが予想されるが、処理コストを抑えることができる。

a) Method I (Batch Algorithm)¹⁾

この手法は、交通状況変化時点を考慮せず、全データにある関係式を当てはめた場合の対数尤度と、変化時点の存在を仮定した場合の対数尤度との尤度比を用いて変化時点を検出する方法である。ここで、対数尤度は式(1)で表される。

$$L(i, j) = m \ln(S^2/m) \quad (1)$$

ここに、時点*i*から*j*までに収集したデータに対して、 $L(i, j)$ は対数尤度、 m はデータ数、 S^2 はデータ平均値に対す

*キーワード: プローブカー, 交通情報, ITS

**正員, 博士(工学), 名古屋大学大学院工学研究科

(名古屋市千種区不老町, TEL: 052-789-3565,

E-mail: miwa@civil.nagoya-u.ac.jp)

***正員, 工修, 株式会社日本航空

****正員, 博士(工学), 名古屋大学大学院工学研究科

*****正員, Ph.D., 名古屋大学大学院環境学研究科

る残差平方和とする。

この手法の変化時点検出手続きを以降に示す。

<Step 1> 基準となる対数尤度を式(1)より算出し、これを L_0 とする。(開始時はデータ数を n とすれば $L(1, n)$)

<Step 2> ある時点 j の前後でデータを2つのセグメントに分割し、各セグメントの対数尤度の和を計算する。この処理を全ての時点 j に対して行い、対数尤度が最小となる時点 k を候補変化時点、対数尤度を L_1 とする(式(2))。

$$L_1 = \min\{L(1, j) + L(j+1, n)\} \quad (2)$$

<Step 3> Step 2で検出した候補変化時点の対数尤度 L_1 を式(3)に代入し、これを満たすならば、先の候補変化時点を第一変化時点とする。さらに、変化時点前後の対数尤度の和 L_1 を L_0 とし、次の変化時点を検出するためにStep 2に戻る。ただし、式(3)を満たさないならば、検出処理を終了する。ここで、 s は予め設定する閾値である。

$$(L_0 - L_1)/L_0 > s \quad (3)$$

b) Method II (Batch Algorithm)

この手法は、変化時点検出手続きはMethod Iとほぼ同じであるが、候補変化時点前後のデータセット間で、リンク旅行時間平均値の差について t 検定を行う。

<Step 1> ある時点 j の前後でデータを2つのセグメントに分割し、各セグメントのリンク旅行時間平均値の差について t 値を算出する。これを全ての時点 j に対して行う。

<Step 2> データ数に基づく自由度を考慮した5%有意水準を満たす候補変化時点が複数存在する場合は、有意水準が最も高い時点を第1変化時点とする。有意な変化時点がない場合は、変化が存在しないと終了する。

<Step 3> すでに探索した変化時点で分割された複数のデータセグメント内でStep 1, 2の処理を実行する。この処理を5%有意水準で検出されなくなるまで実行する。

c) Method III (Incremental Algorithm)

この手法は、Method Iと同様に尤度比により変化時点を検出する手法であるが、データ収集と並行して処理を実行する逐次的な手法である。

<Step 1> 最後に検出された変化時点から最新の収集データまでに対して、Method Iを適用する。

<Step 2> 変化時点を検出しなかった場合は、新たなデータを収集しStep 1に戻る。変化時点を検出した場合は、それ以前のデータを用いないように注意してStep 1に戻る。変化時点検出条件は式(3)と同じであるが、使用するデータ数が少ないため、閾値 s はより大きな値とする。

d) Method IV (Incremental Algorithm)

この手法は、Method IIIがMethod Iを逐次的に実行したように、Method IIを逐次的に実行する手法である。ただし、変化時点検出条件の有意水準は5%のままとする。

e) Method V (Incremental Algorithm)²⁾

以上に示したいずれの手法も、リンク旅行時間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

車両は信号停止の有無によりその分布が多峰性となる場合がある。そこで、より柔軟な手法を考える。この手法は、データ収集と並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データのはずれ値スコアを計算し、変化時点を検出する方法である。

この手法では、確率変数 x が従う確率密度関数として、式(4)、式(5)で表される混合ガウス分布を用いる。

$$p(x|\theta) = \sum_{i=1}^k c_i p(x|\mu_i, \sigma_i^2) \quad (4)$$

$$p(x|\mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right\} \quad (5)$$

ここに、 k は混合する分布の数、 $c_i > 0$ 、 $\sum c_i = 1$ である。また、式(6)を規定する全パラメータをまとめて θ と表す。

このとき、変化時点検出手続きは、以下のStep 1～Step 5をデータ取得と並行して繰り返すものである。

<Step 1> ある最新の(時点 t) データ x_t が得られたとき、以下のSDEMアルゴリズム²⁾により、データが従う確率密度関数を学習する。

Read x_t

for $i = 1, \dots, k$

$$\gamma_i^{(t)} = (1 - \alpha r) \frac{c_i^{(t-1)} p(x_t | \mu_i^{(t-1)}, \sigma_i^{2(t-1)})}{\sum_{i=1}^k c_i^{(t-1)} p(x_t | \mu_i^{(t-1)}, \sigma_i^{2(t-1)})} + \frac{\alpha r}{k} \quad (6.a)$$

$$c_i^{(t)} = (1 - r)c_i^{(t-1)} + r\gamma_i^{(t)} \quad (6.b)$$

$$\bar{\mu}_i^{(t)} = (1 - r)\bar{\mu}_i^{(t-1)} + r\gamma_i^{(t)} \cdot x_t \quad (6.c)$$

$$\mu_i^{(t)} = \bar{\mu}_i^{(t)} / c_i^{(t)} \quad (6.d)$$

$$\bar{\sigma}_i^{2(t)} = (1 - r)\bar{\sigma}_i^{2(t-1)} + r\gamma_i^{(t)} \cdot x_t^2 \quad (6.e)$$

$$\sigma_i^{2(t)} = \bar{\sigma}_i^{2(t)} / c_i^{(t)} - \mu_i^{(t)2} \quad (6.f)$$

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 α 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。

<Step 2> 学習された確率密度関数を用いて式(7)から x_t に対する外れ値スコアを算出する。

$$\text{Score}(x_t) = -\ln p_{t-1}(x_t) \quad (7)$$

これは、 $t-1$ 時点までのデータから学習された確率密度関数に、 t 時点目のデータを入力した場合の対数予測損失を示している。

<Step 3> T移動平均スコア y_t を式(8)により算出する。

$$y_t = \frac{1}{T} \sum_{i=t-T+1}^t \text{Score}(x_i) \quad (8)$$

<Step 4> 得られた移動平均スコア y_t を入力データとして、式(6)により y_t に関する確率密度関数 q_t を学習する。

<Step 5> 式(9)と y_t に関する確率密度関数 q_t を用いてT移

動平均スコアを算出し、これを変化時点スコアとする。

$$\text{Score}(t) = \frac{1}{T} \sum_{i=t-T+1}^t (-\ln q_{t-1}(y_i)) \quad (9)$$

この変化時点スコアが時点 t での変化時点らしさを表し、大きいほど時点 t がより変化時点らしいと解釈できる。

3. シミュレーションデータによる適用可能性検討

(1) シミュレーションデータの概要

ここでは、以上の手法をシミュレーションデータに適用し、各手法の変化時点検出の可能性を確認する。

シミュレーションデータは、名古屋市周辺で収集されたプローブカーデータからDRMリンク旅行時間平均値とその不偏標準偏差の関係式(式(10))を導き、これに基づいて作成した。

$$\text{不偏標準偏差} = 0.469 \times \text{平均リンク旅行時間} \quad (10)$$

各データセットには250のデータを時系列データとして配し、101~150時点のデータが工事等の特異な交通状況の存在により、前後のデータと比べて平均値がZ%ずれているものとする(変化区間)。ここで、このずれ幅を変化幅と呼ぶ。変化幅は平常時のリンク旅行時間平均値に対して、上下に+50% ~ -50% (10%間隔)と設定し、変化区間においてもデータのばらつきは式(10)に従って決定した(図-2)。また、変化区間前後の平均リンク旅行時間は等しいものとし、35secと55secの2パターンを設定した。このようにして計20ケースに対して乱数を変えながら10セットずつシミュレーションデータを作成した。

(2) 適用結果

Batch Algorithmにおける検出の成立条件は、検出された第2変化時点までの少なくとも1つが90~110時点に含まれることとした。一方、Incremental Algorithmについては1~150時点データのみを用い、Method III, IVについては、検出された第2変化時点までの少なくとも1つが90~110時点に含まれることとした。ただし、Method Vについては、その手法の性質を踏まえ、検出された第1~3変化時点までに101~110データ時点が含まれた場合に検出成功とした。なお、Method Iの探索停止条件sは0.5%、Method IIIについては5%と設定した。

表-1に変化時点検出結果を示す。Method I, IIでは、変化区間の変化幅が±30%以上であれば、ある程度の検出が可能となることが分かる。また、Method IIでは、データのばらつきを平均値に比例するように設定したため、変化区間前後の平均旅行時間の大きさによって検出精度が変わらない結果となった。手法間の比較から、Batch Algorithmにおいては、Method IIの方が検出精度が高いことが分かる。逐次的な処理では使用する情報量の低下から検出精度が大きく低下している。ここで、

Method III及びIVについては、変化区間の旅行時間が小さくなった場合でのみ検出に成功しており、変化区間データのばらつきが小さくなったためと考えられる。

Method Vについては、変化区間の変化幅が小さい場合でも検出に成功するケースが見られることから、その適用可能性の高さが伺われる。

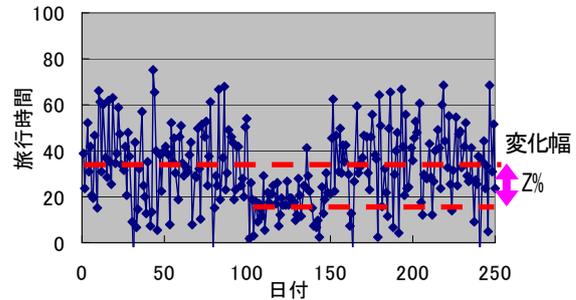


図-2 シミュレーションデータセットの例
(平均: 35sec, 変化区間のずれ: -50%)

表-1 シミュレーションデータによる検出結果

平均リンク 旅行時間	変化幅 Z%	Methodごとの検出精度 (%)				
		I	II	III	IV	V
35sec	+50%	60	70	0	0	20
	+40%	60	70	0	0	20
	+30%	20	30	0	0	20
	+20%	0	0	0	0	20
	+10%	0	0	0	0	20
	-10%	0	0	0	0	20
	-20%	0	10	0	0	20
	-30%	0	30	0	0	20
	-40%	20	70	10	0	20
55sec	+50%	60	70	0	0	20
	+40%	60	70	0	0	20
	+30%	0	30	0	0	20
	+20%	0	0	0	0	20
	+10%	0	0	0	0	20
	-10%	0	0	0	0	20
	-20%	0	10	0	0	20
	-30%	0	30	0	0	10
	-40%	10	70	10	0	20
-50%	70	80	40	10	20	

4. プローブカーデータへの適用

(1) 分析対象エリアの概要

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カーデータである。収集期間は2002年1月~3月、10月~2003年3月までの9ヶ月間であり、2002年1月1日を第1、2003年3月31日を第273と日付換算値を付けた。分析対象エリアは、名古屋市熱田区の路上工事区間とする³⁾が、工事区間で収集されたプローブデータ数が少なかったため、その流入区間を分析対象とする(図-3)。工事区間は国道19号新尾頭交差点~新宮南交差点の約1.5kmであり、期間は2002年3月~2003年2月である。この場合、工事開始日の日付換算値が91、工事終了日は210となる。また、分析区間においては、信号の有無がリンク通過旅行時間に与える影響を考慮し、信号交差点で区切った上下方向の計6区間に対して分析を行った。プローブカー

データ収集量や工事時間帯を考慮し、対象時間帯は13:00~17:00の15分間隔（16時間帯）としている。



図-3 分析対象区間と工事区間

(2) 適用結果

設定した検出成功の条件は、データ取得の日付換算値が76~106の間に検出された場合を成功とし、検出停止閾値はMethod Iで0.5%、Method IIIで30%とした。その他の条件はシミュレーション分析と同様である。

5つの変化時点検出手法の検出精度を表-2に示す。表中の値は、全16ケースに対する検出成功回数である。表より、Batch AlgorithmであるMethod Iの精度が最も高く、次いでMethod IIが高い。また、Incremental AlgorithmであるMethod III~Vは非常に検出精度が低いことが分かる。これは、分析対象区間で実際に収集されたプローブカーデータ数が少ないことが原因であり、逐次的な変化時点検出においては、より多くの（密な）データ量が必要であることが分かる。特に、Method Vは、シミュレーションデータにおいては、Method III, IVより検出精度が高かったが、実際のプローブカーデータには対応できていない。プローブカーのリンク旅行時間は、信号停止の影響で2つのピークを持つ分布となる（図-4）。これに対して、Method Vによる分布の学習過程を確認すると、日付換算91時点（工事開始日）では、学習が十分に行われていないことが分かる（図-5）。すなわち、データの不足から、変化発生時点では分布の学習が十分でないためである。

表-2 プローブカーデータを用いた検出結果

方向	交差点名	Methodごとの検出成功数 (16ケース中)				
		I	II	III	IV	V
南行 (流入)	古渡町→9丁堀	4	4	0	0	0
	9丁堀→金山新橋	4	2	0	0	0
	金山新橋→新尾頭	2	1	0	0	0
北行 (流出)	9丁堀→古渡町	4	2	0	0	0
	金山新橋→9丁堀	4	1	0	0	0
	新尾頭→金山新橋	3	0	1	0	0

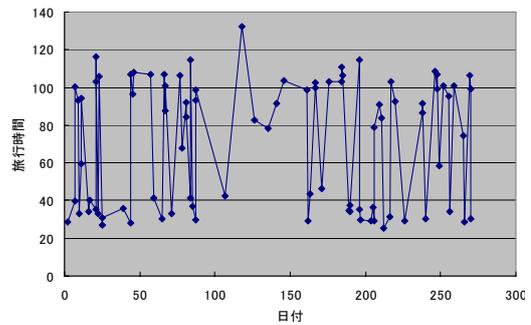


図-4 プローブカーデータのリンク旅行時間分布
(金山新橋交差点→新尾頭交差点 13:00~13:15)

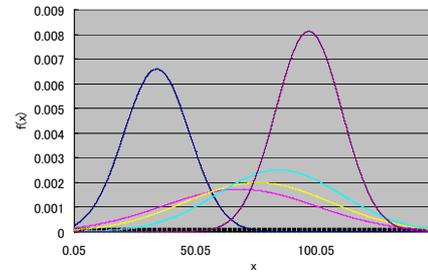
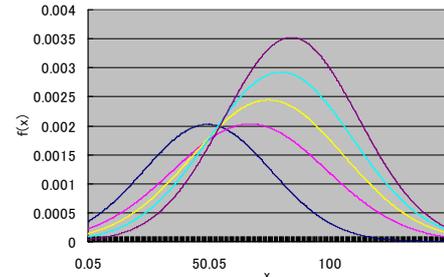


図-5 Method VIにより学習したリンク旅行時間分布
(金山新橋→新尾頭、上：日付換算91、下：日付換算273)

5. おわりに

プローブカーデータからの信頼性の高い交通情報生成を目的として、交通状況が変化する時点を検出するための幾つかの手法を示し、シミュレーションデータ及びプローブカーデータに適用した。この結果より、使用したデータに対しては、蓄積したデータに対して一括処理を行う手法 (Batch Algorithm) の方が、逐次処理を行う手法 (Incremental Algorithm) より精度が高いことが示された。これは、プローブカーデータのようなばらつきの大きなデータからでは、変化時点検出にはより多くの情報量が必要であることが原因であった。今後は、手法の改良とともに、様々な交通状況変化に対して適用を行うことで、より詳細な分析を行う必要がある。

参考文献

- Guralnik, V. and Srivastava, J.: Event detection from time series data, Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp.33-42, 1999
- 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.
- 高松信治ほか：プローブカーデータを活用した路上工事の影響分析, 交通工学, No4, Vol.39, pp.36-41, 2004.