

プローブカーデータを用いた旅行時間の短期予測

杉崎康弘* 藤田貴司** 平林こずえ*

*NECソフト株式会社

**日本電気株式会社

本論文ではプローブカーデータを用いた数十分先までの将来における区間旅行時間を予測処理の方法及び予測精度の評価結果を紹介する。予測方法として、蓄積した過去データとリアルタイムデータを組み合わせた時系列データを用いて、ARモデル(自己回帰)で行う方法を用いた。プローブカーデータは約1500台のタクシーを用いて実施されたインターネットITS実証実験で取得されたデータとした。評価分析については、名古屋の繁華街を通る3つの区間(3.6, 4, 7.6km)での旅行時間情報について予測し、AVI(Automated Vehicle Identification)データを真値(実測値)として予測精度を求めた。評価結果は予測誤差の実測値に対する平均絶対誤差率が十数%であり良好であった。ARモデルを用いる予測方式の有効性が示された。

Short Term Travel Time Coverage estimation for Probe-Car Data

model checking has become an important issue in practical formal verification. Transition traversal coverage focuses on the transition characteristics of CTL

This paper presents the results of our study on taxi operators and calculates travel time using probe-car data. Our method uses the AR model as a prediction function. We analyzed probe-car data from about 1,500 taxis which was collected during a test of the Internet ITS. The performance of our prediction model was validated against the actually observed AVI (Automated Vehicle Identification) data for three different travel distances (length 3.6, 4, and 7.6 km) in Nagoya City. Our results document the effectiveness of our method for short term predictions of travel time.

1 はじめに

旅行時間情報の提供は道路を利用する多くのドライバーから求められている[1]。刻々と変わる道路交通状況が今後どのように変化するのか、旅行時間の予測情報も同様に求められる。その中で、運転中のドライバーにとっては、特に交通状況の変化の激しい突発事象発生時や渋滞の急激な延伸時、解消時などでの、数十分先の将来における交通状況の予測情報は非常に有用である。例えば、動的経路誘導サービスにおいては、未来の交通状況を加味した最適な経路情報の提供が可能となる。このように、予測情報を活用することで道路交通に関するサービスの質

が格段に向上することが期待できる。

また、近年のシステムの高度化に伴い、収集できる交通情報データはプローブカーデータ、AVI(Automated Vehicle Identification)データ、車両感知器データ、トラフィックカウンターデータなど多くの種類がある。そのなかでプローブカーデータは固定インフラ設備を必要とせず比較的、低コストで広範囲の交通情報を収集でき、注目を集めている。また、AVIデータはプローブカーデータとは違い、固定の路線でしかデータを収集できないが、AVIカメラ間については常に大量の旅行時間データが取得できる長所がある。

短期での予測方法については、これまでにいくつ

かの方法が研究されている[2][3]。本論文では、予測方法は[2]で提案されたARモデルを用いた方式を応用した。そして、予測精度の検証は多くの旅行時間データを取得でき真値（実測値）として信頼性が高いAVIデータを用いて、より実証的におこなった。

本論文では、まず利用したプローブカーデータ、データの一次処理の方法について2章で述べ、3章ではARモデルでの予測方法を説明し、4章では真値（実測値）とするAVIデータ及び評価方法について解説し、5章では予測精度結果を示す。

2. プローブカーデータについて

2-1. プローブカーデータ

本論分で使用するプローブカーデータは「インターネットITSプロジェクト」の実証実験により取得されたデータである。実証実験は2002年より名古屋都市圏において1,570台のタクシーをプローブカーとして行われた[4]。データ送信はイベントスキニングにより行われた。イベントは距離周期（300m）、車両発進時、車両停止時の3種類。3種類それぞれの取得されたデータ数はほぼ同等である。収集されるデータの主な項目はデータ収集時刻、車両の位置情報（GPS測位）、（瞬間）速度情報、進行方向情報、タクシーの実車/空車情報である。またGPS車載機の種類があり、その中のType2車載機(655台)はカーナビゲーションシステムが装備されている。

本論文では、カーナビゲーションシステムが装備されているType2車載機(655台)から取得されたデータを分析に用いた。また、空車のタクシーは客待ち等の交通の流れに沿わない運転特性があることも想定されるため、実車データのみを分析に用いるものとした。

2-2. データ処理

予測処理の前処理としてプローブカーデータを用いて2種類の情報を作成する必要がある（図1）。1つは各リンク、時間帯別で過去に取得/蓄積された旅行時間情報の統計情報（蓄積データベース）であり。もう1つは予測処理する時刻の直前における各リンクでの旅行時間情報（リアルタイムリンクデータ）である。本論分では前者を蓄積データベースと呼び、後者をリアルタイムリンクデータと呼ぶ。蓄積データベースは平日及び休日別での典型的なリンク旅行

時間情報を表し、リアルタイムリンクデータは予測処理する時点での特徴的なリンク旅行時間情報を表す。予測処理ではこれらの情報を用いてリンク旅行時間を予測する。

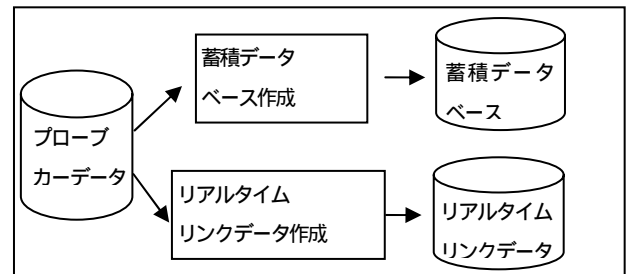


図1 事前のデータ処理 データフロー図

2-3. 蓄積データベースの作成

2002年1～3月、2002年10～2003年3月の9ヶ月分のプローブカーデータから蓄積データベースを作成した。本論分においては平日と休日(土/日/祝日)の2種類を作成した。

以下に手順を示す。

- 1) マップマッチング処理
- 2) 異常データのクレンジング処理
- 3) リンク旅行時間の統計処理
- 4) 蓄積データベースの最適化処理
- 5) 欠損値の補完処理

1) マップマッチング処理（オフライン用）

プローブカーデータはデータ収集時刻情報、（GPS）位置情報、その他の車両の状態情報で構成される車両の軌跡情報である。今回利用するデータにおいては、車両が通過した道路情報は含まれていない。そこで、通過した道路を特定する処理であるマップマッチング処理をおこなった。マップマッチング処理の方法は三輪ら[5]で検討されている。本論文ではこれを参考にして処理をおこなった。

2) 異常データのクレンジング処理

マップマッチング処理（オフライン用）により得られるデータは道路（リンク）ごとの車両の通過情報（リンク旅行時間情報）である。リンク旅行時間情報は各リンクでの通過時刻値とリンク旅行時間の実績値で構成される。この情報には、タクシーの乗客が乗降する時などの個々の車両固有の走行状態や

マップマッチングのエラーも含まれていると考えられる。蓄積データベースは通常時の典型的な交通情報を表すことを目的としており、車両固有の走行状態を含んだデータやマップマッチングのエラーはノイズとなる。そこで[2],[6]と同様に、データマイニングツールとしてNECにて開発されたSmartSifter[7]を用いて異常データの除去処理をおこなった。SmartSifterは混合正規分布を統計モデルとした統計的学習をおこない、その結果を用いて各データの異常値を算出する。平均値から3σ相当以上、外れたデータをノイズとして除去した。

3) リンク旅行時間の統計処理

異常値を取り除いたリンク旅行時間情報を平日、休日別に分け、さらに1日を5分間隔で288の時間帯に分け、それぞれの区分でのリンク旅行時間の統計値(平均値、分散、データ数)を求めた。

4) 蓄積データベースの最適化処理

リンク旅行時間の統計処理において、1日を5分間隔で288の時間帯に分けた。この分割方法は旅行時間の時刻での変動に対して最適であるとは言えない。つまり、深夜等の旅行時間の変動が小さい時間においては、5分間隔の蓄積データベースでは時間帯の幅が細かすぎてリソースに無駄が生じることも考えられる。そこで、[6]と同様に蓄積データベースは記述長最小原理(MDL)[8]を用いて、時間間隔を統計的に最適になるように調整した。

5) 欠損値の補完処理

各リンク、各時間帯でプローブカーが1台も走行しなかった場合、蓄積データベースに欠損値が発生する。この場合には三輪ら[6]と同様に、VICISの旅行時間情報を用いて補完し、これを蓄積データベースとした。

2 - 4 . リアルタイムリンクデータの作成

予測処理する時刻の5分前までのプローブカーデータからリアルタイムリンクデータを作成した。また、本工程はオンラインでのリアルタイム処理で用いることを前提としており、データセンタで収集される大量のプローブカーデータを短時間で処理できることが必要とされる。

以下に手順を示す。

- 1) マップマッチング処理 (オンライン用)
- 2) 異常データのクレンジング処理
- 3) リンク旅行時間の平均処理

1) マップマッチング処理 (オンライン用)

処理の目的は前項のマップマッチング処理(オフライン用)と同様である。しかし、本処理はオンラインでの処理を前提としており、簡略した処理方法である。処理方法は、まずプローブカー1台ごとで連続する2データを用いて車両が通過した道路上の位置を特定し、次いで、道路上の2点のデータ地点間で最短経路探索をおこない、通過した走行経路を特定する。図2に処理のイメージを示す。

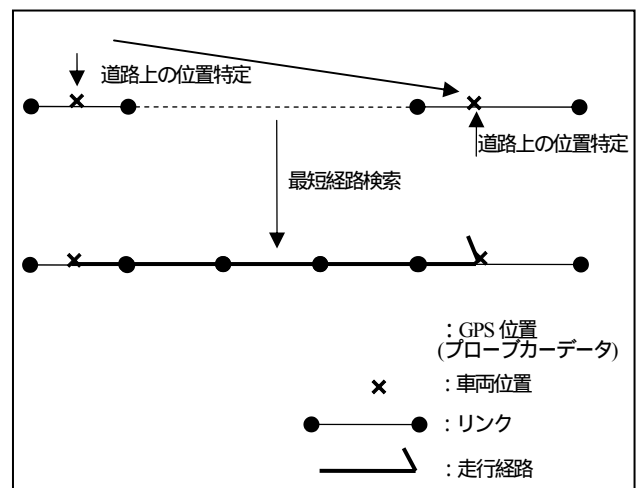


図2 マップマッチング処理(オンライン)イメージ図

2) 異常データのクレンジング処理

マップマッチング処理(オンライン用)で得られるリンク旅行時間情報にはマップマッチング処理のエラーデータや個々の車両固有の走行状態を含んだデータがノイズとして含まれることが考えられる。これを取り除くために、前項の異常データのクレンジング処理と同様にSmartSifterを用いてクレンジング処理を行った。

3) リンク旅行時間の平均処理

異常値を取り除いたリンク旅行時間情報の平均値を求め、これをリアルタイムリンクデータとする。

3. 旅行時間予測の方法

3-1. 予測処理の概要

予測処理は[2]で提案されたARモデルによる方式を用いる。以下ではこの方式をAR予測処理と呼ぶ。AR予測処理の概念はある道路における直前(0~30分)の交通状況は短期的な未来(0~30分)においても同様の傾向が継続されると想定する。つまり、普段の旅行時間と当日の旅行時間を比較して差が大きく、現状の道路状況が普段より混雑している場合には、数十分先までの短期的な将来もある程度、混雑すると想定する。図3でイメージ図を示す。また、直前の交通状況が今後の状況に及ぼす影響の度合いは道路ごとで違ふと考えられる。そこで、道路ごとで影響の度合いを事前に学習処理し、予測時には道路ごとに学習結果を用いて予測処理をおこなう。

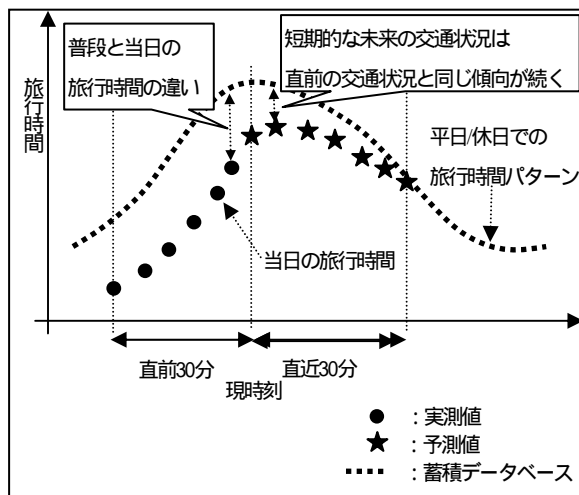


図3 AR予測処理 イメージ図

3-2. AR予測処理

AR予測処理はARモデル(自己回帰モデル)を時系列モデルとして適用して予測する処理である。式(1)にk次のARモデルの定義を示す。ARモデルはある時系列データ x_t について、現在の値から直前のk個の値との間に、線形式が成立すると仮定して、将来の値を予測する手法である。

$$x_t = \sum_{m=1}^k a_m x_{t-m} + \varepsilon_t \quad (1)$$

ここに、 x_t は1次元時系列データ(期待値0を仮定)、 a_m はAR係数、 ε_t は期待値0のガウスノイズである。

このモデルは時系列データ x の定常性を仮定する。一方で、旅行時間情報には24時間や7日間などの周期成分が含まれており、直接に x としては利用できない。そこで、リアルタイムリンクデータから蓄積データベースの値を差し引き、周期成分を取り除き、得られた差分値を x として利用する。また、 x は5分間隔での時系列データとした。このデータをARモデルにより予測処理することの有効性は中田らの研究[2]で確認されている。

実際の処理は学習フェーズと予測フェーズの2つのフェーズに分けて行う。

1) 学習フェーズ

(1)式において、AR係数 a_m 及び、ノイズ ε_t の分散を過去の履歴情報($x_1, x_2, x_3, \dots, x_n$)に基づいて最尤法を用いて推定する。また、本論文では数十分先までの予測を目的としており、 $k=6$ のモデルを構築した。

2) 予測フェーズ

予測処理は以下の(2)式で表される。

$$\hat{x}_t = \sum_{m=1}^k a_m x_{t-m} \quad (2)$$

ここに、 \hat{x}_t は予測値、 x_{t-m} は実測値(期待値0を仮定)、 a_m は学習フェーズで得られたAR係数 a_m ($m=1,2,\dots,6$)である。また、リアルタイムリンクデータが欠損の場合には、 x_{t-m} には過去に取得されたデータから予測する結果を代入する。

3-3. 区間旅行時間の予測

区間旅行時間の予測値を求める際には、区間を構成する全リンクでのリンク旅行時間の予測値の総和で表す。各リンク旅行時間の予測値はリンクへの流入する予想時刻に対応した値とする。つまり、以下の(3)式で表されるタイムスライス値[9]により区間旅行時間の予測値を求める。

リンク数 k で構成される区間の出発時刻 t における区間旅行時間の予測値 $TT(t)$ は

$$TT(t) = \sum_{m=1}^k TT_m(t_m) \quad (3)$$

ここに、 $TT_m(t_m)$ はリンク m における時刻 t_m におけるリンク旅行時間である。 t は区間の出発時刻、 t_m はリンク m への流入する予想時刻である。

4．評価について

4 - 1．評価対象区間

名古屋の繁華街を東西に横切る名古屋長久手線で西向き3区間を対象として、評価をおこなった。区間31は東山から広小路までの7.6km、区間32は東山から今池までの4km、区間21は今池から広小路までの3.6kmである(図4)。特徴としては区間の西側(区間21)が繁華街であり、交通量が多い。また、区間21はリンク数が41、区間32は43で構成される。



図4 評価対象の道路区間

4 - 2．AVIデータ

旅行時間の予測結果を評価するための真値(観測値)としてAVI(Automated Vehicle Identification)カメラで取得したデータ(AVIデータ)を用いる。AVIカメラを東山、今池、広小路に計6箇所、設置し、それぞれの区間を走行する個々の車両の旅行時間を計測した。図5に2004年8月20日のAVIデータを示す。明らかに異常と思われるデータが含まれている。例えば、午前9時頃における旅行時間は多くのデータが約1500(秒)であるのに対して、2500(秒)以上のデータも数件ある。これらの異常値を取り除くために、SmartSifterを用いて、平均値から 2σ 相当以上外れた値を除去した。図6にクレンジング処理後のデータを示す。最終的にはクレンジング処理後のデータの5分間隔ごとの平均値をその区間、その時間での旅行時間の真値として評価をおこなった。

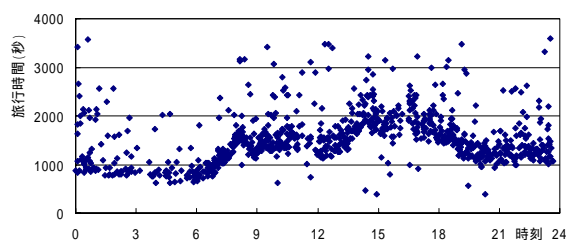


図5 AVIデータ 区間31(8月20日) 1点が1台の車両の旅行時間データを表す。

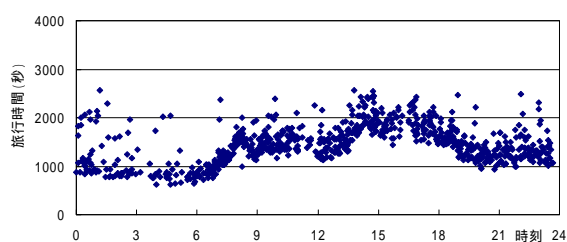


図6 クレンジング処理後 AVIデータ 区間31(8月20日)

4 - 3．精度評価

予測の精度を検証するための指標として、式(4)で表される平均絶対誤差率MAPE(Mean Absolute Percentage Error)値を用いる。

$$MAPE = \frac{\sum_{m=1}^k \left| \frac{x_m - \hat{x}_m}{x_m} \right|}{k} \quad (4)$$

ここに、 \hat{x}_m は m 時点での予測値、 x_m は m 時点での真値(実測値)、 k は予測対象とする時点の数である。

5．旅行時間予測の精度評価

5 - 1．評価結果

以上に述べた方法による予測精度を評価する。図4の3つの区間における、2004年8月3日~8月17日の15日間について、処理をおこない、精度評価をおこなった。結果を表1に示す。

表1より、区間31及び32でのMAPEの全期間での平均値が区間31で約13%、区間32で約12%であり、予測結果は比較的良好である。一方、区間21では約19%である。

図7、図8に誤差値が小さい日の8/7(図7)と大きい日の8/4(図8)の区間21での実績値、蓄積データベース値

および予測値をそれぞれ示す。図7より実測値と蓄積データベースのパターンが比較的似ている典型的な交通状況の日については比較的良好な精度で予測できることが伺える。一方、図8より実測値が蓄積データベースのパターンと違う際に、このずれをAR予測処理が十分に補正できず、誤差が大きいことが伺える。

表2に区間21の8/4,8/7における、リアルタイムリンクデータの総数を示す。区間21を構成するリンク数が41であることから、区間内の約1/4の10リンクのみ、予測の時間間隔（5分間）に1データ以上のリアルタイムリンクデータを使用することが可能であることが伺える。リアルタイムリンクデータの数が少ないこと、すなわちプローブカーデータが少ないことが、AR予測処理が十分に機能しない原因として挙げられる。

表1 旅行時間予測の結果 MAPE値（単位％）

	区間31	区間32	区間21
8/3	14.1	13.4	22.9
8/4	12.3	14.9	21.6
8/5	17.5	12.1	24.9
8/6	12.2	9.5	18.4
8/7	11.9	13.8	16.9
8/8	11.8	11.8	16.7
8/9	11.5	9.0	19.9
8/10	14.7	9.5	23.8
8/11	13.6	11.0	20.1
8/12	13.5	11.7	18.7
8/13	12.5	14.5	16.0
8/14	16.0	19.2	14.9
8/15	12.6	11.5	17.1
8/16	11.2	15.3	17.4
8/17	10.2	9.9	19.1

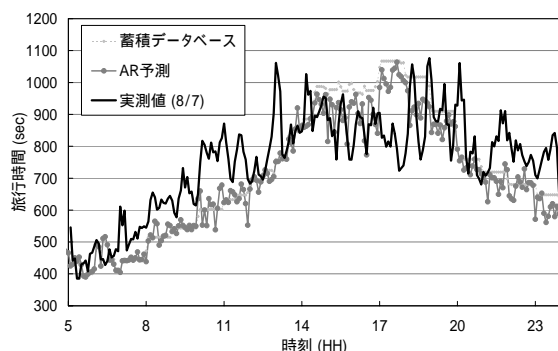


図7 旅行時間予測の結果 区間21（8月7日）

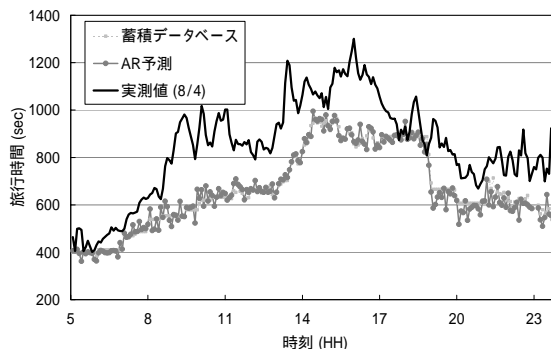


図8 旅行時間予測の結果 区間21（8月4日）

表2 区間21 リアルタイムリンクデータの1日あたりの平均データ数（単位件）

	区間21
8/4	2769
8/7	2565

5 - 2 . 仮想プローブカーデータを用いた評価

プローブカーが十分に走行することを想定した評価をするために、AVIデータを仮想的にプローブカーデータとみなして評価をおこなった。まず、各AVIデータから仮想プローブカーデータとしてリアルタイムリンク情報を作成する。その方法を式(5)で示す。まずクレンジング後の各AVIデータを構成する各リンクの長さに応じて旅行時間を配分し、次いで同じ時間帯におけるリンク旅行時間を平均化する。

$$LT = \frac{\sum_{i=1}^k \left(\frac{l}{L}\right) \times -157}{k} \quad (5)$$

ここでLTはある時間帯におけるリアルタイムリンク情報、 AT_i はある時間帯に*i*番目に取得されたAVIデータ、 l はリンク長、 L はAVI区間長、 k はある時間帯に得られたAVIデータの数。

また、表3にAVIデータのリアルタイムリンクデータへ変換後の1日あたりの平均データ数を示す。区間を構成するリンク数が41であることから、区間内のほぼ全リンクで、予測の時間間隔（5分間）に1データ以上のリアルタイムリンクデータを使用することが可能であることが伺える。

上記で得られるリアルタイムリンク情報を用いて図4の3つの区間で、2004年8月3日～8月17日の15日間について、予測処理をおこない、精度評価をおこな

った。結果を表4に示す。

表4より、各区間でのMAPEの全期間での平均値が区間31で約11%、区間32で約9%、そして区間21では約14%と改善され、予測結果は比較的良好である。

図9、図10に8/7と8/4の区間21での実績値、蓄積データベース値および予測値をそれぞれ示す。これらより、プローブカーデータが十分に得られる状況について、実績値と蓄積データベースのパターンが比較的似ている典型的な交通状況の日(8/7)でも、パターンが違う特異的な交通状況の日(8/4)でも、比較的良好な精度で予測できることが伺える。

表3 区間21 AVIデータのリアルタイムリンクデータへ変換後の1日あたりの平均データ数(単位件)

	区間21
1日あたりの平均データ数	37925

表4 仮想プローブカーデータを用いた旅行時間予測の結果 MAPE値(単位%)

	区間31	区間32	区間21
8/3	13.6	10.4	13.6
8/4	11.6	11.2	12.9
8/5	14.6	9.4	17.1
8/6	10.5	8.9	13.4
8/7	10.0	8.5	15.2
8/8	10.5	9.3	14.1
8/9	10.4	8.8	13.6
8/10	12.4	8.9	16.4
8/11	10.7	9.2	13.6
8/12	11.4	9.8	13.8
8/13	10.3	10.0	13.2
8/14	12.5	10.2	15.2
8/15	10.8	8.4	14.6
8/16	9.8	10.2	14.2
8/17	9.7	8.0	13.7

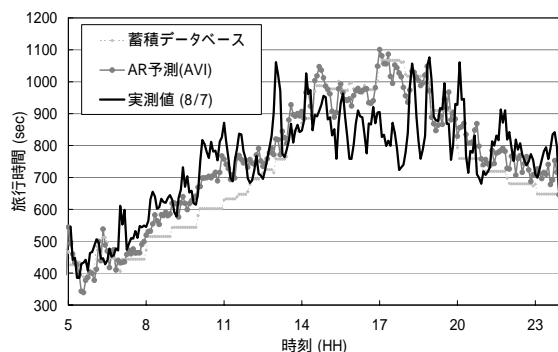


図9 仮想プローブカーデータを用いた旅行時間予測の結果 区間21(8月7日)

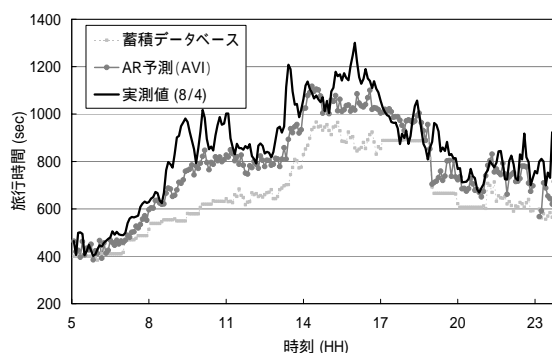


図10 仮想プローブカーデータを用いた旅行時間予測の結果 区間21(8月4日)

6. おわりに

本論文ではプローブカーデータを用いたARモデルによる数十分先までの旅行時間の予測方法を紹介した。また、AVIデータを真値として用いて実証的に精度検証をおこなった。

まず、プローブカーデータを用いてリアルタイムリンク情報を作成して、予測処理および精度検証をおこなった。結果、絶対平均誤差率が区間31で約13%、区間32で約12%、区間21では約19%であった。検証したところ、交通状況が特異な日については精度が上がらないことが示された。その主な原因はリアルタイム情報として利用できるプローブカーデータが少ないことが挙げられた。

そこで、次にプローブカーが十分に走行することを想定した評価をするためにAVIデータを用いてリアルタイムリンク情報を作成して、予測処理および精度検証をおこなった。結果、絶対平均誤差率が区間31で約11%、区間32で約9%、区間21では約14%と精度が向上し、予測結果は比較的良好であった。これらより、ARモデルを用いる予測方式の有効性が示された。

今後の課題を以下に挙げる。プローブカーシステムではある路線について見たとき、旅行時間情報が必ず取得できるわけではなく、取得できるデータ量は一定ではない。そこでプローブカーデータのデータ量と予測精度及び予測日の普段の交通状況との乖離の関係を定量的に調べる必要がある。

また、プローブ情報が少ない路線で且つ交通状況がいつもと違う特異な日については、予測精度の向上が必要である。改善方法としては、プローブ情報

と別の車両感知器等で得られた交通情報を組み合わせて活用することが挙げられる。また交通状況が普段と違う場合には予測旅行時間の値をオフセットするなど、ARモデルとは別の予測方法を用いることが挙げられる。

謝辞

本研究を進めるに際して、AVIデータの取得にご協力いただきましたOriental Consultants Co., Ltdならびにデータ分析のご協力いただきましたP-DRGSコンソーシアム[10]のメンバー、及び関係各位に深く感謝いたします。

参考文献

[1] 割田博, Edward Chung, 竹平誠治, 山田純司: 所要時間情報提供に関する利用者評価調査, 第3回ITSシンポジウム2004 proceedings, pp.107-112, 2004

[2] T. Nakata & J. Takeuchi: Mining Traffic Data from Probe-Car System for Travel Time Prediction, Proc. of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press (KDD2004), 2004.

[3] 堀口良太, 赤羽弘和, 尾高寛信: 累積交通量とアップリンク情報を用いた高速道路旅行時間の短期予測, 第2回ITSシンポジウム2003 proceedings, pp.289-294, 2003

[4] 慶應義塾大学インターネットITS共同研究グループ: インターネットITS研究開発報告書, 平成14年6月

[5] 三輪富生, 境隆晃, 森川高行: プローブカーデータを用いた経路特定手法と旅行時間推定に関する研究, 第2回ITSシンポジウム2003 proceedings, pp.277-282, 2003

[6] 三輪富生, 森川高行, 山本俊行, 藤田貴司, 中田貴之, 竹内純一: プローブカーデータを用いたマルチモード経路案内システム『PRONAVI』の開発と今後の展開, 土木計画学研究・講演集, Vol.31, CD-ROM

[7] K.Yamanishi, J.Takeuchi, G.Williams, and P.Milne: On-line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms, *Data Mining and Knowledge Discovery Journal*, 8(3): 275-300, May 2004.

[8] J.Rissanen: Modeling by shortest data description, *Automatica*, vol.14, pp.465-471, 1978.

[9] 中村英樹, 王鋭: 車両感知器とプローブデータを組み合わせた旅行時間推定手法の実証分析, 第2回ITSシンポジウム2003 proceedings, pp.283-288, 2003

[10] P-DRGSホームページ <http://www.p-drgs.com/>