

路線バスの運行データ分析に基づく到着時刻予測と精度解析

今井 瞳¹ 廣井 慧^{2,3} 河口 信夫^{1,2,3}

概要：本研究では、名古屋市交通局から提供されたバスの位置情報データに対して統計解析を行うことにより、バス到着時刻を予測することを目的とする。従来のバス到着時刻予測は、バス停間での所要時間やバスの接近を予測することに着目した研究が多い。また、得られた結果をどのように評価し、利用者に提示するかといったことに着目した研究は少ない。そこで本研究では、バスが「あと何分でバス停に到着するか」という具体的な到着時刻の予測とその推定精度を提示できるバス到着予測を目指す。到着予測時刻と同時に、運行データの統計解析結果を用いて誤差範囲を利用者に提示する。到着予測時刻がどの程度の推定精度であるか同時に提示することで、利用者の利便性向上を目的としている。目的のバス停に近づくにつれて、その予測精度が向上することが特徴としてあげられる。

Arrival Prediction and Precision Analysis based on Bus Traffic Data

HITOMI IMAI¹ KEI HIROI^{2,3} NOBUO KAWAGUCHI^{1,2,3}

1. 研究背景

市営バスは多くの人に日常的に利用されている。平成 26 年度の国土交通省の調査 [1] によると、名古屋市の 1 年間のバス利用者は延べ約 2 億 8000 万人おり、このことから市営バスは主要交通機関の 1 つであると考えられる。しかし、交通状況や天候などの影響により [2]、市営バスは正確な到着時刻予測が難しい現状にある [3]。また、従来のバス到着時刻予測は、バス停間での所要時間やバスの接近を予測するものが多く、予測結果の評価方法や利用者への提示方法に着目した研究は少ない。

バスロケーションシステム [4] とは、無線通信や GPS などを用いてバスの位置情報を取得し、バス停の表示板やインターネットにおいて情報を提示することで、運行管理を効率的に行い、利用者サービスを向上するシステムである。従来、バスの位置情報はバス停に設置されたアンテナによる無線通信によって取得されるバス停到着・出発情報のみであった。しかし、近年名古屋市では、従来のバス停到着・出発情報に加え、バスに搭載された GPS から 30 秒毎の位置情報が取得可能となった。これにより、リアルタイムに

バスの位置を把握することが可能となり、より推定精度の高いバス到着時刻予測が提示できると考える。

本研究では、名古屋市交通局より、NPO 法人位置情報サービス研究機構 (Lisra) を通じて提供されたバスの位置情報データに対して統計解析を行うことにより、バス到着時刻を予測することを目的とする。従来のバス到着時刻予測は、バス停間での所要時間やバスの接近を予測することに着目した研究が多い。また、得られた結果をどのように評価し、利用者に提示するかといったことに着目した研究は少ない。そこで本研究では、バスが「あと何分でバス停に到着するか」という具体的な到着時刻の予測とその推定精度を提示できるバス到着予測を目指す。到着予測時刻と同時に、運行データの統計解析結果を用いて誤差範囲を利用者に提示する。到着予測時刻がどの程度の推定精度であるかを同時に提示することで、利用者の利便性向上を目的としている。本手法の特徴として、バスが目的地に近づくにつれて、その予測精度が向上することがあげられる。

2. 先行研究

バス到着時刻予測の先行研究として、重回帰分析を用いた手法 [5] やカルマンフィルタを用いた手法 [6] などが挙げられる。以下では、それぞれの手法に関して説明する。

¹ 名古屋大学大学院工学研究科

² 名古屋大学未来社会創造機構

³ NPO 法人位置情報サービス研究機構 (Lisra)

2.1 重回帰分析を用いたバス到着時刻予測

重回帰分析は、予測の対象となる目的変数を、要因として考えられる複数の説明変数を用いて統計的に予測する手法である。Patnaik らによるバスの到着時刻予測を行った研究 [5] では、バス停間の所要時間や距離、乗降者数、ドアの開閉時間等を用いて予測を行っており、高い精度での予測を可能にしている。しかし、重回帰分析は過去の運行データをもとにした静的な予測であり、雨天時やバス停近隣のイベント開催などバス乗客の増加を考慮していない。したがって、このようなリアルタイムに変化する環境に対応し利用者に対してバスの到着時刻を提示するのは難しい。

2.2 カルマンフィルタを用いたバス到着時刻予測

カルマンフィルタは、直前までの情報と現在取得したデータを基に、最も最適なシステムの状態を推定する手法である。Chen らの研究 [6] では、バスがあるバス停区間を通過した際に、バス停区間での所要時間と出発地点からの累積情報を基に、次のバス停区間での所要時間をリアルタイムに予測している。カルマンフィルタは誤差を含んだ情報を処理し、動的な予測が可能であるが、ある特徴のあるバス停区間、例えば他のルートと比べ著しく遅延が生じるバス停が存在した場合、正確な予測を行うことが困難である。

2.3 予測推定精度提示

現在は利用者に対して、主に時刻表に基づいた到着時間を提示している。しかし、この方法では当日の天候、交通状況などによって変化する到着時刻をリアルタイムに提示することは難しい。また、京都市営バスのバス到着案内システム [7] では、各バス停の案内板で接近案内表示システムを用いて、「1つ前/2つ前のバス停に近づいています」、「まもなくきます (Approaching)」といった情報提示を行っている。名古屋市交通局のなごや乗換ナビでは、スマートフォンアプリを通じて、利用者へバスの接近情報を提供している。これらのサービスは、図1の「従来の表示方法」に示すように、バス停を出発したこと、該当のバス停への接近について情報提示するが、バス停とバス停の間のどの位置にあるかを知ることは出来ない。

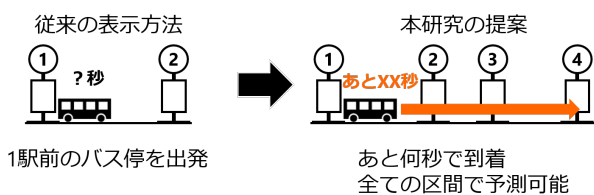


図1 従来手法との比較

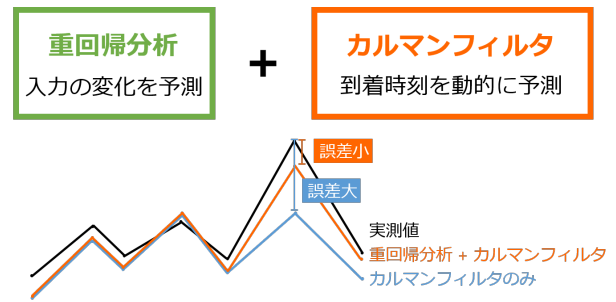


図2 提案手法

3. 重回帰分析とカルマンフィルタを合わせて用いた予測手法

3.1 予測手法の概要

2章で述べたように、重回帰分析を用いたバス到着時刻予測では、過去の運行データから各ルート、バス停間ごとの特色（例えば、あるバス停間では遅れが大きく生じる、あるルートでは雨の日に利用者が増加する等）を算出し、統計学的に予測を行うことが出来るが、動的な予測を行うことには適さない。一方カルマンフィルタを用いた予測では、当日のリアルタイムに取得した累積データを用いて動的な予測を行うことが出来るが、入力の急激な変化に対して適切な対応をすることが難しい。そこで本研究では、図2に示すように、重回帰分析とカルマンフィルタを合わせて用いることで互いの欠点を補完し合う予測方法を提案する。具体的には、名古屋市営バス運行データを基に重回帰分析を行い、各ルート、バス停区間の遅延傾向を統計的に予測する。その後、重回帰分析の結果を初期入力としてカルマンフィルタに入力し、初期入力との誤差をカルマンフィルタによって補填する。本研究の方法を用いることによって、重回帰分析の欠点である動的な予測をカルマンフィルタによって行い、カルマンフィルタの欠点である入力の急激な変化を、重回帰分析で傾向を算出し初期入力とすることで対処することができ、高精度でかつ動的な予測が可能になると考える。

3.2 重回帰分析

3.2.1 予測方法

本研究における重回帰分析では、目的変数 Y をバス停間で生じた相対的な遅延時間、説明変数 X をバス停、曜日、時間帯、遅れ、停止回数、停車時間と仮定した(式1)。

$$Y = a_0 + \sum_{k=1}^N a_k X_k \quad (1)$$

ここでいう遅れとは、ある出発駅からある到着駅の区間の分析を行う際に、出発駅を出発した時点での時刻表からの遅れとし、時刻表の時刻よりも早い場合を負とし、遅い場合を正と定義する。時間帯は、7時までを「Early Morning」、朝の通勤時間帯である8時から10時を「Late Morning」、10

表 1 時間帯

| 時間帯 | 定義 |
|-----------|---------------|
| -7 時 | Early Morning |
| 8 時-10 時 | Late Morning |
| 10 時-13 時 | Early Noon |
| 13 時-17 時 | Late Noon |
| 17 時-19 時 | Evening |
| 19 時- | Night |

表 2 名古屋市営バス運行データ

| |
|----------------------|
| 車載機番号 |
| イベント発生時刻 |
| 緯度・経度 |
| 系統情報 |
| 系統上のバス停 ID |
| バス停上のフラグ |
| イベント種別 (出発, 到着, その他) |

表 3 今回使用したデータ

| ルート ID | ルート名 |
|--------|-----------------------|
| 8415 | [基幹 1] 栄 - 笠寺駅 |
| 8471 | [名駅 1 9] 港区役所 - 名古屋駅 |
| 8784 | [黒川 1 4] 安井町西 - 黒川 |
| 8921 | [平針 1 2] 地下鉄原 - 地下鉄平針 |
| 8939 | [鳴子 1 4] 地下鉄鳴子北 - 大高駅 |
| 8990 | [東巡回] 大曾根 - 茶屋ヶ坂 |
| 9014 | [富田巡回] 戸田 - 戸田荘 |
| 9015 | [富田巡回] 戸田荘 - 戸田 |

時から 13 時までを「Early Noon」、13 時から 17 時までを「Late Noon」、帰宅時間帯である 17 時から 19 時までを「Evening」、19 時以降を「Night」と定義した(表 1)。また、停止回数は速度が時速 5km 未満になった場合を遅れに影響するものとみなし、バス停間で時速 5km 未満になった回数を算出した。停車時間はバス停に到着してから出発するまでの時間とする。

3.2.2 使用データ

名古屋市交通局より提供された 2014 年 12 月 13 日(土)から 12 月 19 日(金)の名古屋市営バス運行データを用いて重回帰分析を行った。提供されたデータには名古屋市内のバス 1030 台、バス停 3784 駅、系統 664 ルートが含まれており、1 日あたり約 220 万桁のデータ量が存在する。バスに搭載された GPS により、従来のバス停到着出発情報に加え、30 秒毎の位置情報が記録されている。その他のデータとして記録される内容は、車載機番号、イベント発生時刻、緯度・経度、系統情報、系統上のバス停 ID、バス停上のフラグ、イベント種別(出発, 到着, その他)である(表 2)。これらのルートの内、今回は以下の表にまとめる 8 ルートについて分析を行った(表 3)。

3.2.3 係数計算

これらのルートの内、先の要素に対して、ルート 8415・[基

表 4 バス停区間

| バス停番号 | バス停区間 | 係数 | p 値 |
|-----------|--------------|--------|-----------|
| バス停 2-3 | 栄-矢場町 | 0 | - |
| バス停 3-4 | 矢場町-丸田町 | -12.8 | 0.02 |
| バス停 4-5 | 丸田町-鶴舞公園前 | -121.4 | <2E-16 * |
| バス停 5-6 | 鶴舞公園前-東郊通三丁目 | 8.9 | 0.07 |
| バス停 6-7 | 東郊通三丁目-高辻 | 25.8 | 6.1E-05 * |
| バス停 7-8 | 高辻-高辻 | -50.4 | <2E-16 * |
| バス停 8-9 | 高辻-雁道 | 30.6 | 1.5E-10 * |
| バス停 9-10 | 雁道-堀田通五丁目 | 11.3 | 0.02 |
| バス停 10-11 | 堀田通五丁目-牛巻 | -19.8 | 6.5E-05 * |
| バス停 11-12 | 牛巻-名鉄堀田 | 16.0 | 0.0009 * |
| バス停 12-13 | 名鉄堀田-地下鉄堀田 | -111.4 | <2E-16 * |
| バス停 13-14 | 地下鉄堀田-呼続大橋 | 4.3 | 0.4 |
| バス停 14-15 | 呼続大橋-千竜通二丁目 | 13.0 | 0.009 |

表 5 その他の要素

| 時間帯 | 係数 | p 値 |
|---------------|-------|----------|
| Early Morning | 0 | - |
| Late Morning | -4.2 | 0.2 |
| Early Noon | -4.2 | 0.2 |
| Late Noon | -2.8 | 0.3 |
| Evening | -6.5 | 0.05 |
| Night | -3.4 | 0.4 |
| 曜日 | 係数 | p 値 |
| Monday | 0 | - |
| Tuesday | 1.4 | 0.7 |
| Wednesday | 0.3 | 0.9 |
| Thursday | 4.5 | 0.2 |
| Friday | 11.0 | 0.001 |
| Saturday | 8.1 | 0.03 |
| Sunday | 6.9 | 0.06 |
| その他 | 係数 | p 値 |
| 遅れ | -1.0 | <2E-16 * |
| 平均停車時間 | 1.0 | <2E-16 * |
| 平均停止回数 | 38.5 | <2E-16 * |
| 切片 | -53.4 | <2E-16 * |

幹 1] 栄 - 笠寺駅について重回帰分析を行った結果を以下の表に示す(表 4-表 5)。表 4-表 5 に示されるように、「栄-矢場町間」、「Early Morning」、「Monday」をそれぞれ基準 0 として係数を算出した。ただし、バス停 1-2:栄-栄間での所要時間は、バス停 1 において運転手による出発時刻記録が正確ではないため、考慮しないものとする。p 値とは、計算結果が統計的に有意かどうかを判断するためのものであり、重回帰分析の場合、「その要素が他の要素に比べて重要ではない」という仮定が正しい確率を表す。すなわち、本研究の場合、p 値が小さいほど遅延に関して重要な要素であることを表している。表 4-表 5 の * は p 値が 0.001 より小さい要素を表し、他の要素に比べて遅延への影響が大きい要素であると考えられる。表 4 より、バス停区間に関しては、大きく遅れるところや大きく遅れを回復する区間など、

表 6 決定係数

| ルート ID | 決定係数 |
|--------|--------|
| 8415 | 0.7907 |
| 8471 | 0.7629 |
| 8784 | 0.6934 |
| 8921 | 0.5838 |
| 8939 | 0.8028 |
| 8990 | 0.7999 |
| 9014 | 0.8006 |
| 9015 | 0.8764 |

各区分毎の傾向の違いが確認できた。表 5 より、時間帯と曜日に関しては、 p 値が 0.001 より小さくなるものではなく、その他のパラメータより遅延への影響が少ないが、曜日の中では金曜日、時間帯の中では「Early Morning」の係数の値が最も大きく、遅延が生じやすいという結果になった。また、遅れ、平均停車時間、平均停止回数に関しては、 p 値は全て 0.001 より小さくなり、遅延に関して重要な要素であることがわかった。それぞれについて具体的にみると、遅れに関しては負の係数として出ているため、1 つ前の時点で時刻表よりも遅ければ早めようと、早ければ遅めようと、時刻表との差を埋めようとする運転手の働きが表れているものと考えられる。平均停車時間に関しては、係数が 1.0 となり、バス停での停車時間がそのまま遅延につながるという結果となった。平均停止回数に関しては、1 回の停止によって、約 40 秒の遅延が生じるという結果となった。

3.2.4 係数計算

決定係数とは、仮定した重回帰モデルによってどれくらい遅延が説明できるかを表す値であり、1.0 に近いほどモデルの精度が高いことを意味する。3.2.3 項における重回帰分析による各ルートの決定係数を以下の表 6 に示す。表 6 に示すように、決定係数はいずれも 0.5 以上となり、強い相関を示した。

3.3 カルマンフィルタ

3.2 節で求めた重回帰分析の結果をもとに、図 3 に示すカルマンフィルタを用いたバス到着時刻予測を行う。カルマンフィルタによるバス到着時刻予測では、まずシステムの初期入力として重回帰モデルによる予測値を入力し、予め遅延が著しく生じる区間等に対応出来るようにしておく。バスが出発した後は、バス停に到着するごとにシステムの状態、及び予測値の更新を行い、バスが終点に到着するまでこれを繰り返す。更新を繰り返すことで、重回帰モデルによる予測値が実測値と異なっていた場合にも修正を行うことができ、またバスが終点に近づくにつれて精度を向上させることが出来る。具体的には、カルマンフィルタでは、時刻 $(k+1)$ におけるシステムの状態は 1 つ前の状態を基に (式 2) で表される状態方程式を用いて求められる。

$$X_{k+1,j} = \Phi_{k+1}x_{k,j} + u_k + Wk, j \quad (2)$$

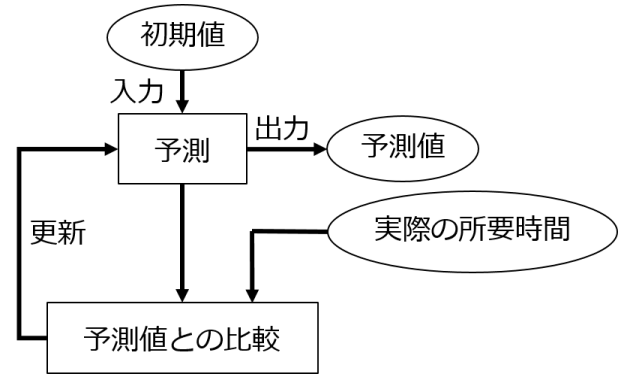


図 3 カルマンフィルタを用いた遅延予測

Φ_{k+1} はシステムの状態遷移に関わる線形モデル、 u_k は状態ベクトル、 Wk, j はノイズを表す。また、観測量 z_k と状態変数 $x_{k,j}$ は (式 3) で表される観測方程式の関係をもつ。

$$z_k = H_k x_{k,j} + v_{k,j} \quad (3)$$

H_k は観測モデル、 $v_{k,j}$ はノイズを表す。本研究では状態変数 $x_{k,j}$ を以下の (式 4) のように時刻表の所要時間 $t_{k,j}$ と実際の所要時間 s_k と定義する。

$$x_{k,j} = (t_{k,j}, s_k) \quad (4)$$

ただし、時刻表の所要時間は任意のバス停 k からバス停 j までの合計値、実際の所要時間は最初のバス停からバス停 k までの合計値とする。

4. 評価

4.1 重回帰モデルに関する評価

4.1.1 評価方法

重回帰モデルの妥当性を確認するため、重回帰モデルによる予測値と 1 週間前の同じルートのデータを用いた予測値の比較を行った。このとき、平均停車回数、平均停止回数は 12 月 15 日のデータを時間帯で平均化した値を用い、遅れについては予測したバス停間で累積した値を用いた。今回使用したルートの内、例としてルート 8415: [基幹 1] 栄 - 笠寺駅について比較を行った結果を以下の図 4-図 5 に示す。縦軸は 12 月 22 日の実測値を、横軸は重回帰モデル、1 週間前のデータによる予測値をとりプロットしたものであり、それぞれの図において予測値と実測値の相関係数が大きくなるほど、予測が正確であると考えられる。図 4-図 5 より、重回帰モデルを用いた予測では相関係数が 0.2424 となり、1 週間前のデータを用いた予測では相関係数は 0.7796 となり、重回帰モデルを用いた予測値の方が、過去のデータを用いて単純な予測を行うよりも精度が高いということが分かる。また、誤差は図 4 が約 80 秒であったのに対し、図 5 では約 43 秒となり、誤差が小さくなったことが確認できた。

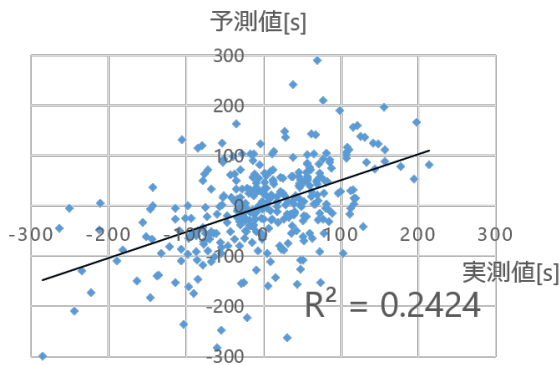


図 4 1週間前のデータを用いた予測値

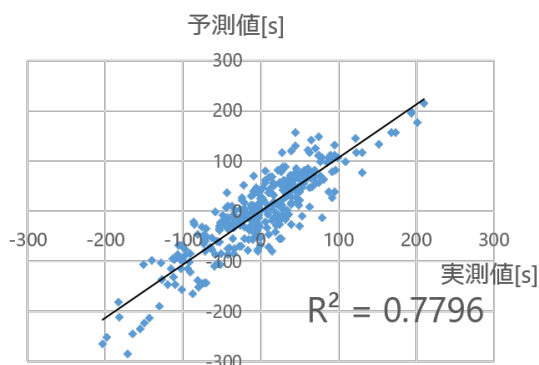


図 5 重回帰モデルによる予測値

4.2 提案手法を用いたバス到着時刻予測に関する評価

4.2.1 評価方法

提案手法の精度向上を確認するため、カルマンフィルタの初期入力時刻を時刻表の値としたものと重回帰モデルによる予測値としたもので比較を行った。カルマンフィルタによる予測結果は、それぞれ RMSE(式 5) を用いて評価した。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5)$$

N はバス停区間の数、 y_i は i 番目の区間の実測値、 \hat{y}_i は i 番目の区間の予測値を表す。RMSE とは、予測値が実測値からどれくらい離れているかを示す評価方法で、0 に近いほど予測が正確であると考えられる。本研究では、12 月 15 日の始発データを用いて、12 月 22 日の始発が各駅バス停に到着する時刻を予測した。重回帰モデルと同様に、ルート 8415:[基幹 1] 栄 - 笠寺駅についてカルマンフィルタによる予測値と運行データによる実測値との差を予測誤差としてまとめたグラフを図 6-図 7 に示す。縦軸に予測誤差、横軸にバス停番号をとっており、予測誤差が 0.0 に近いほど予測結果が良いことを表す。図 6 が初期入力として時刻表の所要時間を入力したもの、図 7 が重回帰モデルから予測される所

表 7 RMSE

| 更新回数 | 時刻表入力 [秒] | 重回帰入力 [秒] |
|------|-----------|-----------|
| 1 回 | 123.9 | 42.8 |
| 2 回 | 97.0 | 53.0 |
| 3 回 | 171.4 | 55.9 |
| 4 回 | 172.8 | 57.3 |
| 5 回 | 190.2 | 69.9 |
| 6 回 | 197.2 | 70.3 |
| 7 回 | 150.8 | 26.5 |
| 8 回 | 121.0 | 20.6 |
| 9 回 | 245.3 | 25.0 |
| 10 回 | 265.4 | 23.5 |
| 11 回 | 278.8 | 31.4 |
| 12 回 | 199.9 | 63.0 |
| 13 回 | 207.8 | 24.2 |
| 14 回 | 173.1 | 50.5 |

要時間を入力したものである。図より時刻表入力に比べ、重回帰モデルの結果を入力したものの方が終点に近づくにつれ予測誤差が 0.0 に近づき、予測結果が修正されていることが読み取れる。また、それぞれの RMSE の値を表 7 に示す。表 7 より、初期入力として時刻表の所要時間を入力したものよりも重回帰モデルから予測される所要時間を入力したものの方が、RMSE の値が小さくなり、予測精度が高いことが確認できた。

4.2.2 考察

重回帰分析によって、比較的高い精度で 1 週間前のデータから所要時間を予測することができた。ルートによって決定係数が著しく悪くなることもなく、本研究における重回帰モデルはバス到着時刻予測に十分利用可能であると考えられる。また、カルマンフィルタによる RMSE の値を、以下の図 8 に示す。図 8 より更新回数が 5 回から 10 回の間では、更新されるにつれて予測精度が向上していることがわかる。しかし、更新回数が 1 回から 5 回、また、11 回以降ではそのような傾向は見られない。本来ならばカルマンフィルタによって更新回数が増えるほど最適状態に近づくため、予測精度が向上するはずであり、不自然な結果となった。

4.3 精度提示方法の検討

提案手法による予測結果を用いて、図 9 のように、利用者が待機するバス停から n 駅前のバス停で行う到着時刻予測の予測精度を算出する。4.2.1 節の図 6 で表される結果を、 n 駅前毎のデータに変換したものを以下図 10 に示す。図 10 より、標準偏差はガウス分布を仮定していることから、約 68.27% の確率で最大誤差 58.6 秒、約 95.45% の確率で最大誤差 117.2 秒以内でバス到着時刻を予測できるといえる。このように、予測時間と共に誤差範囲を提示することで、利用者はより正確にバス到着時刻を把握することができ、行動しやすくなると考えられる (図 11)。

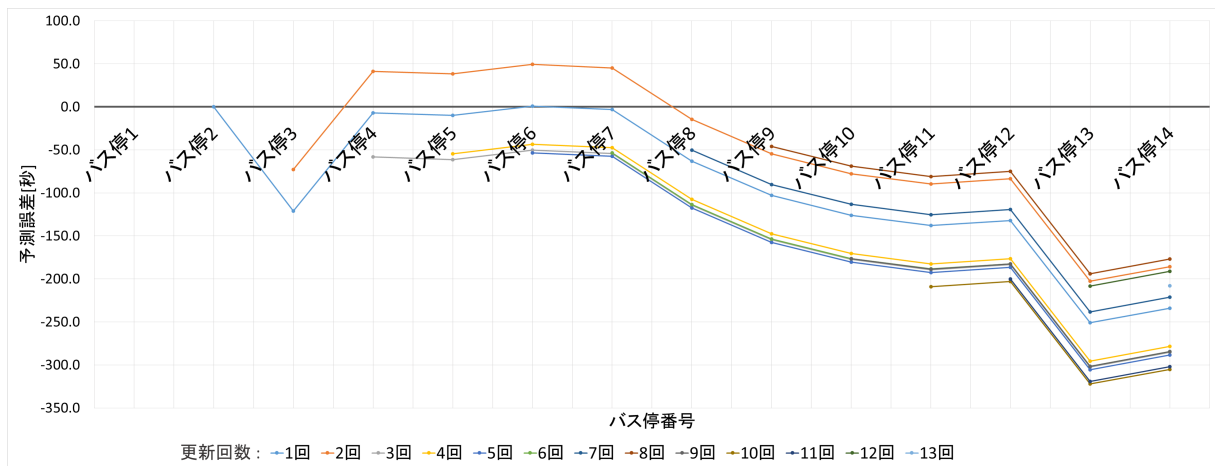


図 6 時刻表入力した時の予測誤差

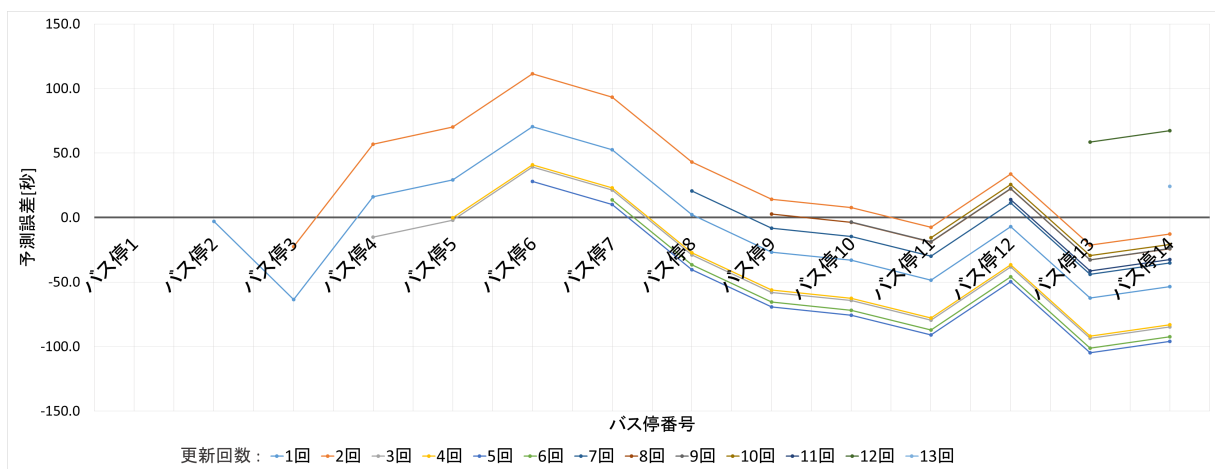


図 7 重回帰入力したときの予測誤差

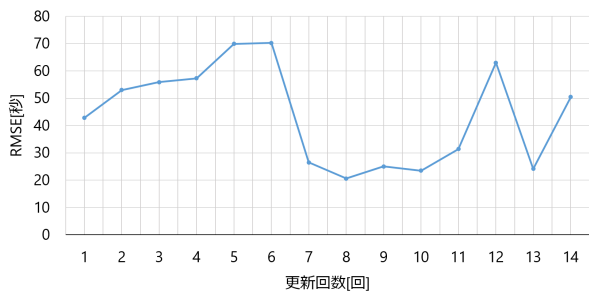


図 8 提案手法による RMSE

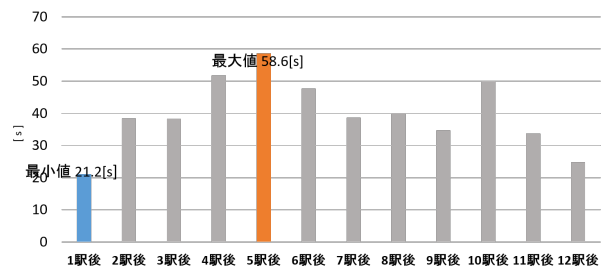


図 10 n 駅前のバス停における予測の標準偏差

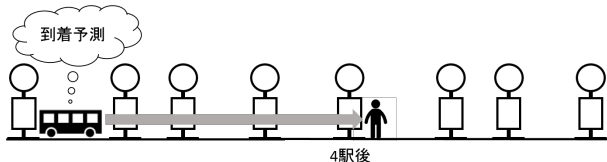


図 9 n 駅前のバス停における予測

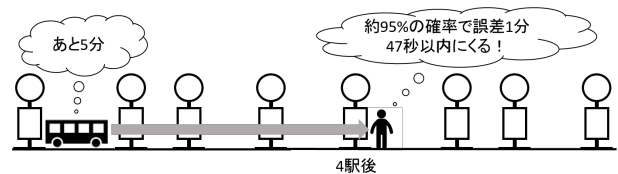


図 11 利用者への提示方法

5. まとめと今後の課題

本研究では、名古屋市営バス運行データに対して、重回帰

分析とカルマンフィルタを合わせて用いるバス到着時刻予測手法を検討した。重回帰モデルによりルート毎の傾向を出発前に算出し、それを初期入力としてカルマンフィルタを利用することで、リアルタイムに予測精度を更新する予

測手法を実現した。今後の課題として、バス停区間の一部でカルマンフィルタによる更新が最適に行われない問題の改善や、今回検証できたルートはごく一部であるため、より多くのルートにおける提案手法の検証が必要であると考えられる。また、利用者がより快適にバスを利用できるような具体的な予測結果提示方法も考えたい。

参考文献

- [1] 統計情報. <http://www.mlit.go.jp/index.html>, 2016.05.11.
- [2] 藤原由美恵, 白石陽. 道路データとバス運行データを用いた路線バスの遅延要因分析. 情報処理学会第 78 回全国大会, 4U-07, 2016.
- [3] Takashi Nagatani. Dynamical transitions to chaotic and periodic motions of two shuttle buses. *JOURNAL OF THE TRANSPORTATION RESEARCH BOARD JANUARY*, 2002.
- [4] 白石陽前川裕一. 乗降者数データと運行実績データを用いたバス到着時刻予測. 情報処理学会第 76 回全国大会, 2V-3, 2014.
- [5] Steven Chien Jayakrishna Patnaik and Athanassios Bladikas. Estimation of bus arrival times using apc data. *Journal of Public Transportation*, Vol. 7, No. 1, pp. 1 – 20, 2004.
- [6] Xiaobo Liu Mei Chen and Jingxin Xia. Dynamic prediction method with schedule recovery impact for bus arrival time. *Transportation Research Record Journal of the Transportation Research Board*, pp. 208 – 217, 2005.
- [7] 京都市営バスが ibeacon で通知、停留所の画面に「まもなく到着」。 http://cloud.watch.impress.co.jp/docs/news/20141104_ghg674245.html, 2016.05.11.