

# ランダムフォレストによるバス走行履歴情報を用いたバス停間所要時間の予測

名城大学 学生会員 ○鷗田一博  
 名城大学 正会員 松本幸正  
 名城大学 非会員 鈴木秀和

## 1. はじめに

バスのサービス改善を目的として、各地でバスロケーションシステム(バスロケ)の導入が進んでいる。このようなシステムの導入を行うことで、定時性の低いバスに対する利用者のストレスが緩和できると考えられる。しかしながら、多くのバスロケは、バスの現在地情報の提供に留まっている。

そこで本研究では、バスロケで到着時刻の予測情報を提供するための第一歩として、過去に蓄積されたバス走行履歴情報を用いて、バス停間の所要時間を予測するモデルの開発を目指す。

## 2. バス停間所要予測時間モデル

### 2.1 ランダムフォレストを用いたモデル

ランダムフォレスト(RF)は、分類や回帰に用いられる機械学習のアルゴリズムの1つである。RFモデルは、ランダムに抽出したデータから決定木を成長させる工程を繰り返すことで構築される。長所としては、過学習が起りにくいこと、データの標準化や正規化の必要がないこと等が挙げられる。

### 2.2 重回帰分析を用いたモデル

重回帰分析は、要因と考えられる複数の説明変数を用いて目的変数を予測する手法であり、広く一般に用いられ、実用的でもある。本研究では、予測結果の比較のために用いる。

## 3. バス停間所要時間予測モデルの構築

本研究では、愛知県日進市のコミュニティバスであるくるりんばすの循環線の全区間を対象とし、所要時間予測を行う。走行情報データの対象期間は2018年6月4日から2018年8月3日までの約2か月である。

本研究における所要時間予測モデルの目的変数は、停車時間を含めた対象バス停区間の所要時間(20秒単位)とした。今回は予測開始バス停を出発してから1区間先、3区間先、5区間先のバス停に到着するまでの3パターンの所要時間の予測を同時に行った。

表-1 は本研究のモデル構築に用いた全ての説明変数の一覧である。説明変数は目的変数に影響を与

表-1 予測モデルの説明変数

運行日・時	時間帯(6分割)
	曜日(7種)
	休日
	五十日
	天候(3種)
バス停情報	予測区間終点バス停の利用者数(多い/普通/少ない)
	途中バス停付近の病院の有無
	予測区間中の利用者の多い/普通/少ないバス停数
	予測区間のロータリー数
	予測区間の駅数
バス停間情報	予測区間の他バス路線乗換可能バス停数
	信号数
	横断歩道数
	計画所要時間(秒)
	計画評定速度(km/h)
	区間長(m)
過去運行情報	左折数
	右折数
過去運行情報	現便予測区間1つ前の 区間実所要時間/計画所要時間(20%単位)
	前便/前日/前週の予測対象区間の 実所要時間/計画所要時間(20%単位)
	現便予測区間1つ前のバス停出発遅延(分)
	前便/前日/前週の予測区間終点バス停到着遅延(分)

える可能性が考えられる運行日・時間情報、バス停情報、バス停間情報、過去の運行情報の計43項目を設定した。網掛けの変数は定性的データであり、それ以外のは定量的データである。

RFでは、分類木による予測を行うモデル(RF分類モデル)と回帰木による予測を行うモデル(RF回帰モデル)の2種類のモデルを作成した。RFの予測モデルの決定木数、特徴量数はRF分類モデルとRF回帰モデルにおいてそれぞれ500、6と500、14とした。

重回帰分析モデルは、ステップワイズで説明変数を増加させながら作成した。

使用したデータの内、92.3%を学習データ、7.7%を予測データとした。

## 4. 予測結果の評価・考察

予測結果の評価は、決定係数、平均二乗誤差(RMSE)、的中率の3指標により行う。RMSEと的

中率においてはバス停の予測区間数別の評価も行う。

図-1は、予測区間数別の予測手法ごとの決定係数を示した図である。この図より、全ての区間数においてRF回帰モデルの決定係数が最も高いことがわかる。RF回帰モデルはRF分類モデルと違い、予測値が連続値であることが要因であると考えられる。

図-2は、予測区間数別の予測手法ごとのRMSE値を示した図である。この図より、全ての区間数においてRF回帰モデルのRMSE値が最も小さいことがわかる。RF回帰モデルは重回帰モデルに比べて、外挿値の予測に優れていることが要因であると考えられる。

図-3は横軸に実測値、縦軸に予測値をとったモデルごとの散布図である。この図より、他のモデルに比べてRF回帰モデルの予測が最も45度線近傍に集中していることがわかる。

予測値の的中率を実測値との誤差±0分、±1分ごとに算出して、比較する。図-4は予測区間数別の予測手法ごとの的中率を示したものである。この図より、±1分の精度ではどの区間数においてもRF回帰モデルが最も高い数値を示しているが、重回帰モデルと比べて最大3.8%の差に収まっていることが見て取れる。しかし、重回帰モデルは説明変数に1つでも欠損値がある場合に予測値を算出できないことから、データ欠損が見られるバスロケへの導入には適さないと考えられる。

今回の予測結果の評価・比較から、予測する区間数が多くなるほど予測精度が高くなることがわかった。予測区間数が大きくなるとその区間中に時間調整が可能なバス停が多くなり、実所要時間のばらつきが小さくなることが一つの要因であると考えられる。しかしながら、RF回帰モデルでは、1区間予測が±0分での的中率が最も高く、バスロケによって予測到着時刻の情報を提供する際に、出発するごとに次のバス停への到着時刻を予測することが望ましいと考えられる。

5. おわりに

本研究では、実際のバス走行履歴情報を用いて3つの所要時間予測モデルを構築し、モデルの種類と区間数ごとの予測結果にどのような差異があるのかを決定係数、RMSE、的中率により把握した。結果として、RF回帰モデルが他モデルよりも優れてい

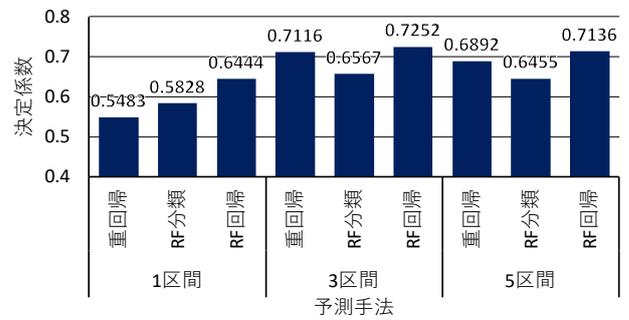


図-1 予測手法・予測区間数別の決定係数

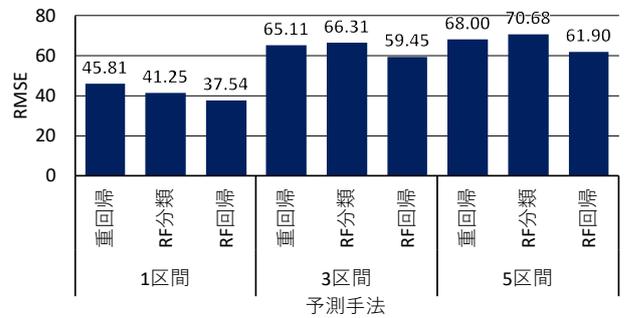


図-2 予測手法・予測区間数別のRMSE値

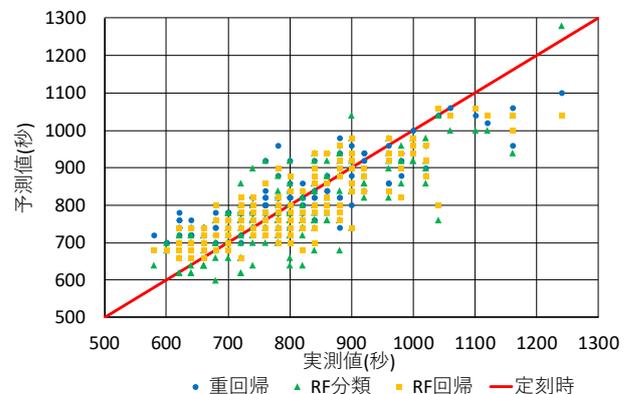


図-3 5区間予測時の予測値と実測値の散布図

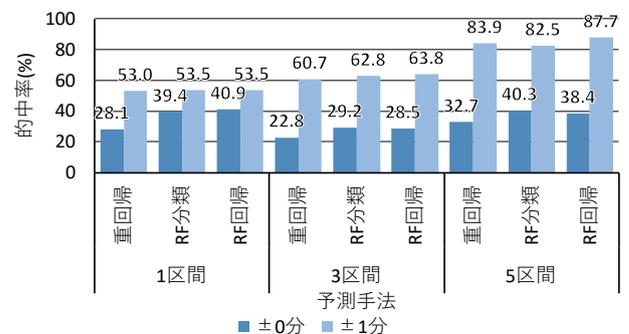


図-4 予測手法・予測区間数別の実測値との的中率

ることがわかり、5区間予測で87.7%が誤差±1分以内に収まった。

今後は、実際のバスロケで所要時間予測システムを用いることを考慮し、リアルタイムデータに対応させていく必要がある。