

# Random Forest 法によるバス遅延要因を考慮した到着時刻の精度向上の検討

## Study on Precision Improvement of Predicting Bus Arrival Time

### Focusing on Delay Factors using Random Forest Method

指導教授 轟 朝幸

5086 寺田 敦

## 1. はじめに

近年、バスの輸送需要が減少傾向にある。バス事業者は利用促進策として利便性の向上のために、バスロケーションシステム（以下、バスロケ）を導入している。バスロケとは、バスの接近やバス停の到着予測時刻の情報を、携帯端末等を通して利用者に提供する仕組みである。事業者による予測到着時刻の提供は、通過停留所の実績時刻に先行バスの実績所要時間を加算するという簡易的なシステムが多い。そのため、正確な予測到着時間を利用者に提供できているとは言い難い。

## 2. 既存研究と本研究の位置づけ

轟らりは、国際興業バスの大 01 系統の下り路線を対象として、機械学習モデル手法の一つであるニューラルネットワークを適用した。遅延時間の予測値と実測値の誤差が 3 分程度の精度で予測モデルの構築を行うことができた。しかし、ニューラルネットワークは予測のプロセスがブラックボックスで要因の影響関係の解釈が難しく、遅延要因となる変数の重要度が分からないという欠点を有する。そのため、変数の影響度を客観的なデータから把握し、精度を向上させることが困難である。

そこで本研究では、遅延要因の向上影響を表現できる遅延時間の予測モデルを構築し、精度に資する重要な変数を明確化する目的とする。遅延要因の具体的な把握によって予測モデルの精緻化が可能となり、利用者の提供情報に対する満足度向上やバス事業者による運行計画等のサービスレベル向上が期待できる。

## 3. 対象路線と分析手法

### 3. 1 対象路線

本研究での対象路線は、国際興業株式会社（国際興業バス）運行する大宮駅東口停留所から浦和美園駅西口停留所間の大 01 系統を対象路線とする。

### 3. 2 分析手法

本研究では、機械学習の中でも可読性が高く、精度の高い予測が可能である Random forest 法を用いて分析を行う。Random forest 法は、分類や回帰に多く用いられる手法であり、学習データから特徴を学習し、決定木と

呼ばれる樹木状の構造を複数作ることで、予測結果を多数決で最終予測する手法である。Random forest 法の特徴として、説明変数の目的変数に対する影響度を視覚化できるため、複雑なブラックボックスモデルを可読性の高い解釈可能なモデルで表現することができる。

## 4. 運行データの基礎集計による現状把握

2014 年 4 月から 2016 年 3 月までの運行等のデータを用いて基礎集計を行った。乗車人数では大宮駅東口停留所、降車人数では浦和美園駅西口停留所がそれぞれ最も多くなった。各停留所の遅延時間では、全ての停留所で遅延していることが明らかになったが、特に根木輪停留所での遅延時間が最大であった。



図－1 路線図概略

## 5. 遅延時間予測モデルの構築

### 5. 1 遅延予測の活用シーン

本研究では、表－1 に示した 2 種類の活用シーンを想定した予測モデルの構築を行う。具体的には、パターン①は大宮駅東口停留所の出発直後にいくつかの到着停留所の遅延時間を予測する場合である。パターン②は、途中の各停留所を通過後に浦和美園駅西口停留所の遅延時間を予測する場合である。それぞれリアルタイムの予測モデルの構築を行い、変数の重要度を明らかにする。

表－1 予測対象

	出発停留所(m)	予測対象(m)
パターン①	・大宮駅東口	・大宮区役所(0.3) ・堀の内橋(1.6) ・南中野(2.7) ・庚申塚(3.7) ・根木輪(5.9) ・三崎台(6.5) ・浦和美園駅西口(12.0)
パターン②	・南中野(2.7) ・根木輪(5.9) ・三崎台(6.5)	・浦和美園駅西口(12.0)

※( )は大宮駅東口停留所からの距離

## 5. 2 モデルに使用する変数

2014 年 4 月から 2015 年 3 月までのデータを学習させ、2015 年 4 月から 2016 年 3 月までの遅延時間の予測を行う。モデルに採用する説明変数を表 2 に示す。道路混雑は Google map の過去の統計に基づくデータを用いる。発車停留所遅延時間などの変数はリアルタイムのデータを用いる。

表 2 モデルに使用する変数

質的変数		量的変数
降雨ダミー	道路混雑	発車停留所遅延時間
休日ダミー	発車時刻	停留所乗車人数
イベントダミー	月	前便遅延時間

## 5. 3 分析結果

図 2 は浦和美園西口停留所の出発直後の Random forest 法による遅延時間の予測値と、実際の実績値とを比較したものである。決定係数は 0.50 とやや低い値となった。実遅延時間と予測遅延時間の的中率は、±3 分以内が 54.5%，±1 分以内が 24.2%であった。決定係数及び的中率が低くなった原因として、エラーデータが分析サンプルから除外されたため、学習データが少なくなったが原因と考えられる。

図 3 は、大宮駅東口停留所出発後の、浦和美園駅西口停留所での遅延時間の予測における、変数の重要度を示したものである。重要度を示す平均増加誤差は、始発遅れが最も高く、次に発車時刻と道路混雑の重要度が高いことが分かった。これにより、時間帯別の道路混雑状況の違いが遅延に影響していることが考えられる。

予測時点を変更したモデルでは、予測タイミングが浦和美園駅西口停留所に近づくにつれて、決定係数及び的中率が高くなった。変数の重要度は、始発遅延及び発車時刻、道路混雑がともに高いことが明らかになり、遅延の要因となっていることが考えられる。

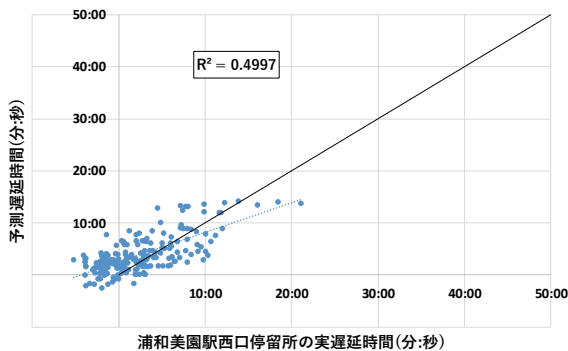


図 2 浦和美園駅の推定状況

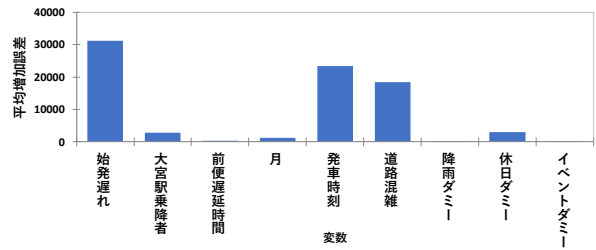


図 3 浦和美園駅の予測における変数の重要度

図 4 に、大宮駅出発後の各停留所の遅延時間の予測モデルの変数の重要度を示す。予測地点を出発地から遠ざけていくにつれ、発車時刻と道路混雑の重要度が高くなることが分かった。これにより、発車停留所から距離が長くなるほど、時間帯別の道路混雑状況が影響していることが考えられる。始発遅れに関しては、いずれの停留所においても最も重要度の高い変数であったが、浦和美園駅西口停留所付近では、重要度がやや下がっていることから、時間調整停留所である三崎停留所にて遅延時間が緩和されたことが影響していると考えられる。

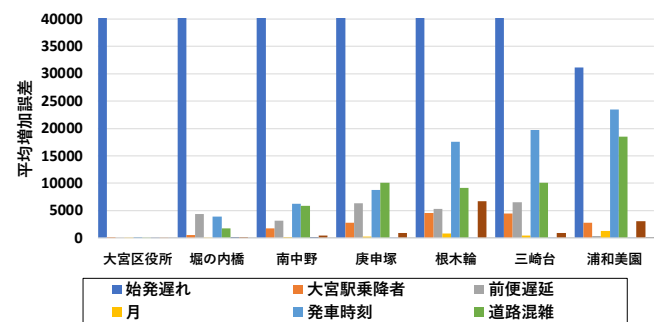


図 4 各バス停の遅延予測における変数の重要度

## 6. おわりに

本研究では、Random forest 法を用いたことで、予測モデルの可読化を行うことができた。モデルより変数の重要度が明らかになり、バスの遅延時間に始発遅れと時間帯別の道路状況が大きく影響していることが分かった。

今後の課題として、道路交通センサスデータなどのより詳細な変数加えるなどして、より精度の高い遅延時間予測モデルの構築を検討していく必要がある。

## 参考文献

- 1) 轟朝幸, 川崎智也, 野村大智, 横関敬裕: ニューラルネットワークを用いた路線バスの遅延時間予測, 交通工学論文集, Vol.3, No.2, pp.202-207, 2017.