

利用者の乗降データを用いたバス乗客人数予測モデルの評価

青地美桜^{1,a)} 白石陽^{2,b)}

概要：公共交通機関の一つとして、路線バスが挙げられるが、バス利用者は路線バスの混雑を事前に把握することが困難である。混雑しているバスを事前に把握することができれば、バス利用者は混雑していないバスを選択し、乗車することが可能となる。混雑状況の予測方法の一つとして、乗客人数に基づく方法がある。既存の手法として、カメラを用いて乗客人数を推定する手法があるが、この手法では、乗客が密集すると正確に推定できない問題がある。そこで本研究では、交通系 IC カードの乗降データを用いてバスの乗客人数を予測する手法を提案する。交通系 IC カードのデータには、乗降停留所、乗降時刻、路線、バスの車両番号等が記録されている。これらのデータを用いることで、指定した路線の停留所間の乗客人数が算出できる。提案手法では、時間帯や天候の変化に応じて乗客人数が変動することを考慮して、各時間帯の乗客人数を目的変数とした予測モデルを作成し、予測精度を評価する。また、乗客人数予測モデル作成に適している手法を検討するために、別の機械学習アルゴリズムを用いて乗客人数予測モデルを作成し、乗客人数の予測を行ってその精度の評価を行う。

キーワード：交通系 IC カード、混雑、公共交通、路線バス、乗客人数予測

1. はじめに

代表的な公共交通機関の一つとして、路線バスが挙げられる。バス利用者が路線バスを快適に利用するには、路線バスの遅延状況と混雑状況の把握が必要である。2016年に内閣府が行った世論調査によると、利用者が鉄道やバスに感じる不満として、「遅延する」(17.7%)と「車内が混雑している」(15.7%)が挙げられている[1]。遅延状況に関して、既存のバスロケーションシステムがあり、バス利用者はバスロケーションシステムを利用することで、遅延状況を考慮してバスを選択することが可能である。しかし、既存のバスロケーションシステムでは混雑状況を考慮してバスを選択することができない。また、地方の路線バスは1時間ごとのバス本数が少ない。混雑しているバスが到着しても、利用者は次のバスへ乗車するという選択をすることができず、バス利用者は、混雑しているバスへ乗車せざるを得ない場合がある。ここで、バスの混雑状況は車内の乗客人数が多いほど混雑していることになる。そこで、バス利用者に事前にバスの乗客人数を提示することで、バス利用者は混雑していないバスを選択することが可能になり、バス内で着席できる可能性が高まる。また、着席できない場合でも、別の乗客と密着するほど混雑しているバスを避けることができるため、バス利用時の快適性が高くなると考える。よって、乗客人数を予測し、利用者へ提示することで、利用者はバス内の混雑状況を把握することができ、バス乗車時の快適性の向上につながると考える。

乗客人数を予測するために、学習データとして乗客人数を使用する。乗客人数は、各停留所で乗車した人数、降車した人数を使用して求めることが可能である。そこで、本

研究では、バスの運賃支払いに使用される交通系 IC カードに着目する。交通系 IC カードは、利用時に乗降時刻、乗降停留所等の乗降データを記録する。乗降データからカード利用者が乗降した停留所および時刻がわかる。よって、ある停留所で乗降した人数を把握できるため、直前の停留所までの乗客人数に乗車人数を加算し、降車人数を減算すると、路線バスに乗車した乗客人数を算出することができる。算出した乗客人数や、該当する停留所、停留所を通過した日時等を学習させることで、乗車予定のバスの乗客人数の予測が可能になると考える。交通系 IC カードの一つである Suica は、2019年3月31日時点で7,587万枚発行されており、前年度より約500万枚増加している[2]。今後さらに、普及率は高くなると考えられるため、乗客人数の予測に活用することは妥当であると考えられる。

そこで、本研究では、バス利用者の快適性を向上させるために、交通系 IC カードの乗降データを使用して、バス内の乗客人数を予測する手法を提案する。著者らの先行研究[3]では、バスの運行時間帯と天候を考慮し乗客人数予測モデルの評価を行ってきた。しかし、乗客人数予測モデルを作成する際に使用した天候データは、降雨の有無データのみであった。降雨の有無データのみでは雨の強さなどを考慮することができず、台風などの悪天候で減少する乗客人数を考慮することができないと考える。本稿では、予測精度を向上させるために、降水量や風速などの新たな説明変数を追加し、乗客人数予測モデルの評価を行う。また、使用するデータ量を増加させ、乗客人数予測モデルの評価を行う。さらに、乗客人数予測モデル作成に適している手法を検討するために、予測手法を変更して、乗客人数予測モデルを作成し、予測精度の評価を行う。

2. 関連研究

本研究では、乗客人数を予測するために必要な乗客人数を、各停留所の乗降車数を用いて算出する。そこで、関連

1 公立はこだて未来大学大学院 システム情報科学研究科
Graduate School of Systems Information Science, Future University Hakodate
2 公立はこだて未来大学 システム情報科学部
School of Systems Information Science, Future University Hakodate
a) g2120001@fun.ac.jp
b) siraisi@fun.ac.jp

研究として、2.1 節では、乗降車数を推定する研究について述べる。2.2 節では、公共交通機関の混雑状況を予測する研究について述べる。

2.1 乗降車数を推定する研究

中島らはドライブレコーダに記録された映像を用いている[4]。この研究では、背景差分法を用いて動体を検出し、それらの輪郭検出を行い、動画フレーム間の乗客の動きから乗降車数を推定している。提案手法の評価を行った結果、バスの扉を開けた際に差し込む光の変化を誤って認識してしまうことが原因となり、推定精度が低い結果となった。差し込む光は天候によって変化するため、天候が原因で推定精度が低くなってしまふ可能性がある。そのため、カメラを用いて推定を行う場合、天候を考慮する必要があると考えられる。

山田らは、路線バスの乗降口に測域センサを設置して乗降車数の推定を行っている[5]。測域センサとは、周囲の物体との距離を広範囲で正確に把握することができるセンサである。測域センサの計測データはセンサから見た方位角と距離で表すことのできる計測対象の位置情報である。実際に運行しているバスに測域センサを設置し実証実験を行った結果、高い精度で乗降車数を推定することができた。しかし、バスの乗降口付近で多数の乗客が密集してしまった場合、密接している乗客を1人と判定してしまい、推定結果との誤差が大きくなってしまふ問題がある。混雑するバス内の乗客人数は誤差が1~2人以上である場合、バスに乗車することが難しくなる可能性がある。そのため、バスの乗降付近で乗客が密集した場合にもより正確にバス内の乗客人数を把握する必要があると考える。

これらの研究より、バスに直接機器を設置したものを使用して乗降車数を推定した場合、バスの状況により誤差が生じてしまい、十分な精度が得られない可能性が考えられる。

2.2 公共交通機関の混雑状況を予測する研究

前川らは、鉄道に乗車している利用者の携帯端末で受信される Bluetooth の RSSI (Received Signal Strength Indicator) 特性を利用し、混雑状況を推定している[6]。この研究では、端末間で観測される RSSI から、端末が同一車両に存在する確率（以下、同一車両確率）と、端末間が混雑している確率（以下、混雑確率）を算出する。次に、同一車両確率および一部の端末の乗車車両情報を用いて乗客が所有している全端末の車両内での位置を推定し、その推定結果と端末間の混雑確率を用いて、車両ごとの混雑を推定する。乗車車両情報とは、その端末が存在する車両の情報である。提案手法の評価を行った結果、車両ごとの混雑の有無を F 値 0.75 で推定できることが示されている。しかし、乗客が所持している端末の位置を正確に推定することができな

ければ、車内の混雑度合いを把握することが困難である可能性がある。

高松らは、バス内に設置した BLE (Bluetooth Low Energy) 端末から得られる RSSI の減衰特性と、音圧の減衰特性を利用し、混雑状況を推定している[7]。この研究では、混雑状況推定手法として BLE を用いた手法と音を用いた手法の2つを利用する。BLE を用いた手法では、バス内の特定の場所に BLE 端末を設置し、受信端末でそれぞれの BLE 端末から得られる RSSI の減衰を特徴量として利用する。音を用いた手法では、バスのピンポン音を音源とし、バス内に設置したマイクまでの人の遮蔽による音の減衰を特徴として利用する。2つの提案手法でそれぞれ混雑状況を推定した結果、BLE を用いた手法では、乗客の立ち位置を変更しても分類可能であることが示された。音を用いた手法では、詳細な分類は困難だが、マイク近傍が無人であるか、そうでないかを高確率で推定可能であることが示された。しかし、BLE 端末およびマイクの設置箇所によって推定結果が変化してしまう可能性がある。特に、BLE 端末と受信端末との位置関係が重要であるため、様々な環境に対応することが困難である問題がある。

佐藤らは、過去のバス利用者数と天候、人口データを用いて、予測モデルを作成し、翌日のバス利用者数の予測を行っている[8]。天気予報の気温、降水確率、停留所周辺の世帯数、年齢別人口、乗車数、降車数、停留所を通過した人数を学習データとして、予測モデルを作成し、予測を行っている。しかし、この研究では、バスの運行時間帯を考慮していないため、通勤、通学時間帯等の特定の時間帯で増加するバス利用者数の予測に対して十分な精度が得られていないという問題がある。

3. 提案手法

本章では、3.1 節で本研究における研究目的について述べる。次に 3.2 節で研究における研究課題とそのアプローチについて述べ、3.3 節で提案手法の流れについて述べる。最後に、3.4 節で本研究で使用するデータについて述べる。

3.1 研究目的

本研究の目的は、路線バス内の乗客人数予測を行い、利用者の快適性を向上させることである。本稿では、乗客人数予測を行うための予測モデルの予測精度向上を目的とし、作成した予測モデルの評価を行う。ここで、本研究における乗客人数とは、バスに乗車している利用者の人数のことである。また、バスへ乗車する人数、バスから降車する人数を総称して乗降車数とする。

3.2 研究課題とアプローチ

本研究における研究課題を次の5つに示す。

- a) 乗客人数予測に使用する乗降データの検討
- b) 時間帯および天候による乗客人数の変化の考慮
- c) 乗客人数に有効な説明変数の検討
- d) 乗客人数予測に使用するデータ量の検討
- e) 予測モデル作成に利用するアルゴリズムの検討

これらの課題に対するアプローチを以下に述べる。

課題 a に対するアプローチとして、バス乗車時に利用する交通系 IC カードの乗降データに加えて、バスの運行実績データを用いる。バス内の乗客人数予測を行うためには、バスが停留所を通過した時刻と、そのときの乗客人数のデータが必要である。乗客人数は交通系 IC カードの乗降データから算出することが可能であるが、バスが停留所を通過した時刻は乗降データのみでは抽出することができない。停留所で乗降した利用者が存在する場合、交通系 IC カードの乗降データから乗降時刻を抽出できるが、利用者の乗降が発生しない場合は停留所を通過した時刻を把握することができない。そこで、停留所を通過した時刻を算出するために、バスの運行実績データを利用する。バスの運行実績データには、バスが停留所を実際に通過した時刻のデータや、計画時刻のデータが記録されている。バスの運行実績データと乗降データ中の乗降時刻を関連付けることで、乗降車が発生していない場合でもその停留所通過時点の時刻と乗客人数を把握することができる。

課題 b に対するアプローチとして、本研究ではバスが運行している時間帯と、悪天候かを考慮して予測を行う。バスの乗客人数は、通勤・通学が原因で朝に増加することがあり、加えて、悪天候の場合さらに増加する可能性がある。佐藤らは、バス利用者の予測を行っているが、通学時間帯におけるバス利用者の増加を予測することができていない[8]。時間帯と天候を考慮することで、乗客人数の予測精度を向上させることができると考える。したがって、本研究では、機械学習を用いて、バスが運行している日時、時間帯、天候、そのバスの乗客人数を学習させ、予測モデルを作成し、乗客人数予測を行う。

課題 c に対するアプローチとして、説明変数に、交通系 IC カードの乗降データ、時間帯データ、降雨の有無、降水量データ、風速データを使用する。予測モデルの作成に用いた目的変数と説明変数の一覧を表 1 に示す。目的変数は各停留所におけるバス内の乗客人数とした。説明変数は表 1 に示す 7 種類とした。表 1 の説明変数の詳細を述べる。停留所番号、運行日時、曜日を用いることで、各停留所のある運行時間の乗客人数を予測することができる。ここで、停留所番号は、停留所を一意に特定できる番号のことである。時間帯データは、定義した時間区分にあてはまる時間帯カテゴリに分類したものである。時間帯データを用いることで、バスの運行時間帯による乗客人数の違いを考慮することができる。天候データは、降雨の有無、降水量、風

速のデータを用いる。降雨の有無は気象庁が公開しているデータから、降雨の有無を判別し、2 値のデータを作成する。降水量、風速データは気象庁が記録している数値を用いる。降雨の有無、降水量、風速データを用いることで、天候の悪化の程度がわかり、悪天候時の乗客人数の増減を考慮することができる。

表 1 目的変数と説明変数の一覧

対応変数	使用データ	データの内容
目的変数	乗客人数 (人)	ある停留所におけるバス内の乗客人数
説明変数	運行日時	バスが運行した日時
	曜日	バスが運行した曜日
	停留所番号	停留所に割り振られている番号
	時間帯	定義した時間区分にあてはまる時間帯カテゴリ
	降雨の有無	運行時の降雨の有無
	降水量	運行時の降水量
	風速	運行時の風速

課題 d に対するアプローチとして、乗客人数予測に使用するデータ量を増加させる。本研究では、乗客人数予測モデルを作成するために様々なデータを用いているが、データが少ない場合、予測モデルを作成する際に学習不足となってしまう、予測精度が低くなってしまいう可能性がある。著者らは、これまでも乗客人数予測モデルを作成し評価を行ってきており[3]、データ量は 2019 年 4 月～8 月までの 5 ヶ月分のデータのみを使用していた。よって、本稿では雪が降っていない 2019 年 9 月、10 月のデータを新たに追加し、データ量を増加させることで乗客人数予測モデルの予測精度が向上するかを検討する。

課題 e に対するアプローチとして、ランダムフォレスト、一般化線形モデル、一般化加法モデルを使用し、予測モデルを作成し比較する。ランダムフォレストは決定木を弱学習器とするアンサンブル学習アルゴリズムである。決定木を複数用いて、それぞれの決定木の結果を補うことで高い予測性能を得ることができる[9]。これランダムフォレストの利点の一つとして、未知のデータに対する汎化性能が高いことが挙げられる。バスの乗客人数を予測する際には、イベントの発生、天候の悪化等で乗客人数が通常の状態から増減することが考えられる。そのため、ランダムフォレストを用いることで、乗客人数を予測した際の誤差を小さくすることができる可能性があると考えられる。一般化線形モデルとは、目的変数が正規分布やポアソン分布等の確率分布に従っている場合に適用することができる[10]。また、一

一般化加法モデルは、目的変数と説明変数が非線形な関係でも適用できるように、一般化線形モデルを拡張したものである[11]。角野らは奈良県で運行する路線バスの IC カードデータを用いて、バス利用者数が正規分布に従うかの分析を行っている[12]。結果として、バス利用者が正規分布に従う便、対数正規分布に従う便、どちらにも従う便が 62 便中 59 便、どちらにも従わない便が 62 便中 3 便となっている。よって、バス利用者数は正規分布に従っていることが示唆されている。そのため、今回予測する乗客人数も正規分布に従っている可能性が高いため、一般化線形モデルを用いることで予測精度が高くなる可能性がある。また、イベントや悪天候が原因となり、目的変数と説明変数が非線形な関係となる可能性もあるため、一般化加法モデルを用いて予測モデルを作成し、予測精度を比較する必要があると考える。

3.3 提案手法の流れ

以下に本研究の提案手法の流れを示す。

- ① データ収集
- ② データの前処理
- ③ 予測モデルの作成
- ④ 乗客人数の予測

まず、バスの乗客人数を予測するために使用するデータを収集し、整理する (①)。

次に収集したデータで前処理が必要なデータは前処理を行う (②)。前処理が必要なデータとしては、交通系 IC カードの乗降データ、天候データが挙げられる。交通系 IC カードの乗降データの前処理として、運行実績データと運行時刻を関連付け、乗降データと運行実績データから乗降車数を求める。求めた乗降車数から乗客人数を求め、その乗客人数を予測モデルの作成に使用する。停留所 N でバスへ乗車する人数を x_N 、バスから降車する人数を y_N 、停留所 N の一つ前の停留所である停留所 $N-1$ までの乗客人数を P_{N-1} とするとき、停留所 N における乗客人数 P_N を求める式を式 (1) に示す。

$$P_N = P_{N-1} - y_N + x_N \quad (1)$$

天候データの前処理として、天候データは「晴れ」、「曇り」、「雨」等の天気の情報記録されているため、それらのデータを雨が降っているか、降っていないかの 2 値のデータを作成する。

次に作成したデータを用いて、乗客人数予測モデルを作成する (③)。

最後に、作成した予測モデルを用いて乗客人数の予測を行う (④)。

3.4 本研究で使用するデータ

本研究で使用するデータとして、交通系 IC カードの乗降データ、運行実績データ、天候データについて述べる。3.4.1 項で交通系 IC カードの乗降データ、3.4.2 項で運行実績データ、3.4.3 項で天候データに関して詳しく述べる。

3.4.1 交通系 IC カードの乗降データ

交通系 IC カードの乗降データは、カード利用者がバス乗降車時に交通系 IC カードを使用すると記録されるデータである。表 2 に実際に記録されるデータの一部を示す。

表 2 交通系 IC カードの乗降データの内容

データ名	内容
乗車日時	バスに乗車した日時
乗車停留所番号	乗車した停留所に割り振られている番号
乗車停留所名	乗車した停留所名
降車日時	バスから降車した日時
降車停留所番号	降車した停留所に割り振られている番号
降車停留所名	降車した停留所名
車両番号	利用したバス車両の番号
路線番号	利用した路線に割り振られている番号
路線名	利用した路線名

乗車日時および降車日時は、カード利用者がバスに乗車およびバスから降車した日付と時刻が記録されている。乗車停留所番号および降車停留所番号は、カード利用者がバスに乗車およびバスから降車した停留所の番号が記録されている。停留所番号とは、停留所を一意に特定することができるように、バス会社が定義した番号のことである。乗車停留所名および降車停留所名は、カード利用者がバスへ乗車および降車した停留所の名称が記録されている。車両番号は、カード利用者が乗車したバスの車両番号が記録されている。路線番号は、カード利用者が乗車した路線番号が記録されている。路線番号とは、バス路線を一意に特定することができるように、バス会社が定義している番号のことである。路線名は、カード利用者が乗車したバスの路線名が記録されている。

3.4.2 運行実績データ

運行実績データは、各停留所での実際のバス運行情報をまとめたデータである。表 3 に実際に記録されるデータを示す。

路線番号は、運行しているバスの路線番号が記録される。上下区分は、バスが運行している方向が記録される。本研

究で用いたデータでは、バスの終点方向が駅方向の場合「往」、その逆向きが「復」と記録されている。通過順序は、バス運行時に通過した停留所を何番目に通過したかという情報が記録されている。停留所名は、バスが通過した停留所名が記録されている。計画通過時刻は、バス会社がバス運行時に計画した停留所を通過する時刻である。記録されている時刻は時刻表に記載されている時刻である。実績通過時刻は、バス運行時に実際に停留所を通過した時刻が記録されている。

表 3 運行実績データの内容

データ名	内容
路線番号	バスが運行している路線の番号
上下区分	バスが運行している方向
通過順序	バスがその停留所を通過した順序
停留所名	バスが通過した停留所の名前
計画通過時刻	時刻表上のバスが停留所を通過する時刻
実績通過時刻	バスが実際に停留所を通過した時刻
車両番号	運行しているバスの車両の番号

3.4.3 天候データ

悪天候時の通勤・通学手段を変更する場合がある。例えば、晴れている日に自転車で通勤・通学をしているが、雨が降った日はバスに乗り換えるような場合である。このように、悪天候時に通勤・通学手段を変更する場合があるため、悪天候時にはバスの乗客人数が増加することが考えられる。また、台風等の外出することが困難な天候の場合、乗客人数が減少することが考えられる。前川らは、乗降客数を用いてバスの到着時刻予測を行うために、まず、バスの乗降客数データを分析している[13]。分析結果として、悪天候時に乗降人数が変化することが示唆されている。よって、本研究では、バス利用時の天候がバスの乗客人数に影響すると考え、バス運行時の天候データを使用する。天候データには「晴れ」や「曇り」、「雨」等の細かい天気の情報、気温、降水量、降雪量、風速等のデータが記録されている。本研究では、記録されているデータの内、天気の情報、降水量、風速を用いる。

4. 実験および考察

本章では、まず 4.1 節で説明変数、データ量を変更した予測モデルの評価を行う。4.2 節では予測モデルを複数のアルゴリズムを用いて作成し、予測モデルの精度評価を行う。

4.1 説明変数、データ量を変更した予測モデルの評価

本実験では、説明変数を追加し、データ量を増加させた

結果、予測精度が向上するかを調査するために、予測モデルを作成し、評価を行う。本実験では、説明変数を追加した乗客人数予測モデルを 4 つ、データ量を増加させた乗客人数予測モデル 1 つを作成する。予測モデルの作成には、一般的な予測手法であるランダムフォレストを用いる。実験環境と実験に用いたデータを次項以降に示す。

4.1.1 実験環境

本実験では、統計解析ソフト R を用い、予測モデルの作成を行った。本実験の実験環境を表 4 に示す。

表 4 実験環境

ハードウェア (PC)	
OS	Windows 10 Home 64bit
CPU	Intel(R) Core i5-1035G7
RAM	8.00 GB
プログラム	
R 言語	Ver.3.6.1
実行環境	RGui (64bit) ver.3.6.1

4.1.2 実験に用いたデータ

本項では、実験に用いたデータについて述べる。本研究では、函館バス株式会社から提供を受けた交通系 IC カードの乗降データおよび運行実績データを用いる。具体的には、2019 年 4 月から 10 月までの期間で実際に運行していた 15 路線分のデータを利用する。本実験では、15 路線の中で函館市内の複数の主要な停留所を通過する路線である路線 A のデータを用いる。なお、2019 年 4 月第 1 週は天候が雪であったため、除外した。また、バスの運行実績データが欠損していた日も除外した。対象便は第 1 便 (6 時~7 時台) から最終便である第 16 便 (20 時~21 時台) までである。表 5 に実験に用いたデータの一覧を示す。

表 5 実験に用いたバスの運行データの一覧

項目	内容
対象路線	路線 A
対象期間	2019 年 4 月から 2019 年 10 月
対象便	第 1 便から第 16 便 (最終便)
総データ件数	99,376 件

天候データは気象庁が公開しているデータ[14]を用いる。天候データ内にある天気データは 3 時間毎に記録されているため、天気が記録されていない時間が存在する。本実験では、天気が記録されていない時刻のデータを直前に記録された天気データを用いて補間する。例えば、6 時に天気が記録されている場合、7 時と 8 時は 6 時の天気データを用いる。表 6 に実験に用いた天候データの一覧を示す。

表 6 実験に用いた天候データの一覧

項目	内容
対象地点	函館
対象期間	2019年4月から10月 6時から21時のデータ

4.1.3 実験で用いる説明変数

本実験では、3.2節に示した表1に示す説明変数を用いる。停留所番号は、函館バス株式会社が設定している停留所番号を用いた。時間帯データは、3.4.2項で述べたバスの運行実績データにある実績通過時刻データを用いて、指定した時間区分の時間帯に分類する。本実験では、対象とした路線バスの運行時間である、6:00から21:59までの時間を16分割し、カテゴリを分けたデータを用いる。降雨の有無データは、気象庁が公開している天候データ内の天気データを用いて、降雨の有無を判別し、2値化したデータを用いる。降水量データ、風速データは気象庁が公開している数値を用いる。

4.1.4 作成する乗客人数予測モデル

本実験では、説明変数の追加、データ量の増加で予測精度が向上するかを調査するために、乗客人数予測モデルを複数作成し比較する。作成するモデルを表7に示す。降水量、風速以外の説明変数は表1に示したものである。また、データ量を増加させるモデルは、表7に示したモデルAからモデルDで最も精度が高かったモデルを用い、2019年4～10月のデータを用いる。

表 7 作成する乗客人数予測モデル

モデル	説明変数	使用データ期間
モデル A	降水量なし，風速なし	2019年4～8月
モデル B	降水量あり，風速なし	2019年4～8月
モデル C	降水量なし，風速あり	2019年4～8月
モデル D	降水量あり，風速あり	2019年4～8月

4.1.5 予測モデルの評価方法

本研究では、二乗平均平方根誤差 (Root Mean Square Error, RMSE) を用いて乗客人数予測モデルの評価を行う。RMSEは数値がゼロに近いほど予測精度が高いことを示す。本実験では10-分割交差検証を行い、RMSEを算出した。RMSEの値は予測モデルを用いて予測した乗客人数と実際の乗客人数の誤差の大きさを示している。

RMSEを求める数式を式(2)に示す。式(2)の n はデータ量、 f_i は予測値、 y_i は実測値である。

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (f_i - y_i)^2} \quad (2)$$

4.1.6 説明変数、データ量を変更した予測モデルの評価結果

乗客人数予測モデルを作成し、その予測モデルの精度評価を行った結果を述べる。表8に各予測モデルのRMSEを示す。

表 8 説明変数を変更した予測モデルの精度評価結果

予測モデル	RMSE (人)
モデル A	3.86
モデル B	3.57
モデル C	3.59
モデル D	3.59

表 9 データ量を変更した予測モデルの精度評価結果

予測モデル	RMSE (人)
モデル B	3.57
データ量を増加したモデル B	3.22

4.1.7 考察

表8より、降水量データを追加したモデルBが、予測精度が高いことがわかる。これより、降水量データを追加することで乗客人数をより正確に予測できることが示唆された。これは降雨の有無のみの場合、雨の強さを把握することができなかったためであると考えられる。降水量データを用いることでどの程度雨が降っていたかを考慮することができるため、そのデータを説明変数に追加することで、乗客人数の変化をより考慮することができるようになったと考えられる。また、データ量の増加により予測モデルの精度が向上した理由として、2019年4月～8月の5ヶ月分のデータでは不十分であったと考えられる。本実験では2019年4月～10月の7ヶ月分のデータを使用した。今後もデータ量を増加させることで予測精度が向上する可能性があるため、引き続きデータ量を増加させていく必要があると考えられる。ここでRMSEの値に注目すると、すべての予測モデルで3～4の間となっている。乗客人数を予測する際、バスが混雑していない場合を考えると予測誤差が4人程度であれば影響がないと考える。しかし、バスが混雑している場合、バスに快適に乗車するのが困難になってしまう可能性があるため、今回作成した予測モデルの精度としては不十分であると考えられる。

4.2 乗客人数予測モデル作成方法の評価

本実験では、予測モデル作成に適している手法を検討することを目的とし、一般的な予測手法であるランダムフォレスト、一般化線形モデル、一般化加法モデルを用いて予測モデルを作成し、予測精度を比較する。実験環境は4.1.1項、実験に用いたデータは4.1.2項、実験で用いる説明変

数は 4.1.3 項、予測モデルの評価方法は 4.1.5 項と同様である。

4.2.1 乗客人数予測モデル作成方法の評価結果

ランダムフォレスト、一般化線形モデル、一般化加法モデルを使用し、乗客人数予測モデルを作成し、その予測モデルの評価を行った結果を述べる。なお、ランダムフォレストを使用して作成されたモデルは表 9 のデータ量を増加させたモデル B と同様である。

表 10 乗客人数予測モデルの精度評価結果

使用した手法	RMSE (人)
ランダムフォレスト	3.22
一般化線形モデル	3.57
一般化加法モデル	4.81

4.2.2 考察

表 10 より、ランダムフォレストを用いて予測モデルを作成した場合、最も精度が高いことがわかる。これは説明変数に連続的な数値を使用しているものが少ないからであると考えられる。今回の説明変数では停留所番号、時間帯、曜日、降雨の有無が数値ではなくカテゴリ変数となっているため、ランダムフォレストを使用し決定木を作成するときに分岐を作ることが容易であるからだと考えられる。今後、数値データの説明変数を追加した場合でも、カテゴリ変数が多いため、ランダムフォレストを用いて作成した乗客人数予測モデルを使用した場合、予測精度が高い可能性があると考えられる。

また、一つ前の決定木を次の決定木で修正するように学習を進める勾配ブースティング回帰木という手法がある。今回の実験で、ランダムフォレストを使用した場合に最も精度が高くなったため、今後は勾配ブースティング回帰木を使用した場合と比較し、予測手法を再検討する必要があると考えられる。

5. まとめ

本研究では、バス利用者が路線バスの混雑を把握し、混雑していないバスを選択することをできるようにするために、交通系 IC カードデータの乗降データ、バスの運行実績データ、天候データを用いて乗車予定バスの乗客人数を予測する手法を提案した。提案手法では、バスの乗客人数を予測するために使用するデータを収集し、収集したデータの前処理を行う。前処理を行ったデータを用いて予測モデルを作成し、その予測モデルを用いて乗客人数予測を行う。まず、説明変数、データ量を変更した予測モデルを作成し、予測精度の評価を行った。10-分割交差検証を行った結果、

RMSE の値は、降水量データを説明変数として追加し、データ量を増加させた予測モデルが 3.22 と最も良い精度となった。これより、説明変数として降水量データが予測精度の向上に寄与し、データ量を増加させることで学習不足が改善され、より正確に予測できることが示唆された。次に、予測手法を変更して乗客人数を予測するモデルを作成し、評価を行った。10-分割交差検証を行った結果、RMSE の値はランダムフォレストを用いて作成した予測モデルが 3.22 となった。これより、乗客人数予測モデルはランダムフォレストを用いて作成することで予測精度が向上することが示唆された。

今後は、さらに予測精度を向上させるために必要なデータ量の再検討を行う。また、乗客人数は、地域の祭りなどのイベントでも大幅に変化するため、イベントの考慮を行うことを検討している。さらに、現在は一つの路線のみに対して予測モデルを作成しているが、バス利用者は様々な路線を使用するため、複数の路線のデータを使用し、予測モデルの作成、評価を行うことを検討している。また、ランダムフォレストを用いて予測モデルを作成するとともに、同じく決定木を用いて予測する勾配ブースティング回帰木を使用して予測モデルを作成し、予測手法を再検討する。

謝辞 本研究の実験に用いた交通系 IC カードデータおよび運行実績データは函館バス株式会社の協力によるものである。ここに深く感謝の意を表する。

参考文献

- [1] “公共交通に関する世論調査”，内閣府，
<https://survey.gov-online.go.jp/h28/h28-kotsu/index.html> (参照 2020-01-22)
- [2] “JR 東日本会社要覧，IT・Suica 事業”，JR 東日本，
<https://www.jreast.co.jp/youran/>. (参照 2020-01-07).
- [3] 青地美桜，白石陽，“交通系 IC カードの乗降データを用いた路線バスの乗客人数予測手法の提案”，情報処理学会第 82 回全国大会講演論文集，No.1，pp.353-354 (2020)
- [4] 中島颯人，新井イスマイル，藤川和利，“バス車載ドライブレコーダを用いた背景差分方による乗降客数推定手法の評価”，情報処理学会マルチメディア，分散，協調とモバイルシンポジウム 2018 論文集，Vol.2018，pp43-48 (2018).
- [5] 山田遊馬，廣森聡仁，山口弘純，東野輝夫，“測域センサを利用した高精度な路線バス乗降計測システム”，情報処理学会論文誌，Vol.60，No.3，pp.934-944 (2019).
- [6] 前川勇樹，内山彰，山口弘純，東野輝夫，“鉄道における Bluetooth RSSI 特性を用いた乗車車両および混雑の推定手法”，情報処理学会論文誌，Vol.55，No.6，pp.1614-1624 (2014).
- [7] 高松将也，伊達伸之輔，岩本健嗣，松本三千人，“CroW: バス車内における混雑状況推定システム”，情報処理学会マルチメディア通信と分散処理研究報告(DPS)，No.14，pp.1-7 (2015).
- [8] 佐藤孝大，大場みち子，“バスの到着時刻予測モデルの開発と移動手段提案システムの検討”，情報処理学会研究報告情報システムと社会環境(IS)，No.7，pp.1-7 (2018).
- [9] 波部斉，“ランダムフォレスト”，情報処理学会研究報告コンピュータビジョンとイメージメディア(CVIM)，Vol.2012-CVIM-182，No.31，pp.1-8 (2012).
- [10] 荘島 宏二郎，宇佐美 慧，橋本 貴充，登藤 直弥，高野 慶輔，

“統計モデルの違いを理解する”, 教育心理学年報, Vol.57,
pp.302-308 (2018).

- [11] 宮田敏, “適応型モデル選択基準を用いた一般化加法モデルの推定”, 日本統計学会誌, Vol.39, No.2, pp.369-391 (2010).
- [12] 角野惇, 猪井博登, 山室良徳, 田中徳人, 館秀央, 土井健司, “路線バス事業者における交通系 IC カードデータの利用意向と路線バスの利用変動分析”, 第 53 回土木計画学研究発表会講演集, pp.702-708 (2016)
- [13] 前川裕一, 林夏美, 牧野友哉, 白石陽, “バス到着時刻予測におけるバス運行所要時間データと乗降客数データの活用”, 情報処理学会マルチメディア通信と分散処理ワークショップ論文集, Vol.2013, No.6, pp.165-171 (2013).
- [14] “過去の気象データ・ダウンロード”, 国土交通省 気象庁, <https://www.data.jma.go.jp/gmd/risk/obsdl/>. (参照 2019-12-09)