COVID-19状況下における バス運行状況および運行所要時間予測精度の変化

川谷 卓哉[1,a)]　佐藤 祐大[1,b)]　峯 恒憲[2,c)]

概要：新型コロナウイルス感染症（COVID-19）感染拡大防止に係る措置により，2020 年の 4 月〜6 月は人の移動が大きく縮小され，道路交通量も減少した．このことは，路線バスの運行状況にも何らかの影響を与えるはずである．本論文では，2019 年と 2020 年のそれぞれ 4 月〜6 月における路線バスプローブデータを採取し，統計的な差異を比較した．その結果，運行所要時間（移動時間）の分布には有意差があり，中央値は小さくなっていた．また，2019 年の情報を用いて，2020 年のバスの移動時間を予測したときの予測精度は低下したが，学習データを多くすると精度が改善した．

1. はじめに

バスの運行所要時間（移動時間）や遅延時間を予測することは，バスのサービス水準を向上し，利用客の利便性を向上させるために欠かせない．正確な予測を実現するため，統計的手法や機械学習を利用した，さまざまな予測モデルが提案されている．

しかし，これらの予測モデルの構築，および検証に用いられているデータは 1 週間程度と比較的短期間であることが多く，それらのデータには，沿線でのイベントなど，突発的な交通環境の変化が含まれていないことがほとんどだと考えられる．また，長期的な交通環境の変化，たとえば新規道路の開通や，商業施設の開業，あるいは交通環境を一変させるほどの社会的事象の発生を含むようなデータはまれであろう．最後に挙げた社会的事象のひとつ，それが新型コロナウイルス感染症（COVID-19）である．

新型コロナウイルス感染症（COVID-19）の影響により，2020 年の 4 月から 6 月の期間は，人の移動が大きく縮小された．その結果として，路線バスの利用客数は対前年比で 3 割減〜半減と大幅に減少する結果となり [1]，道路交通量も減少した [2]．また，本研究で調査対象としたバス路線・昭和自動車九州大学線においては，運行本数が最大で 63%削減された．このことは，路線バスの運行状況，具体的にはバス運行時の遅延にも何らかの影響を与えるはずである．

著者らは，先行研究 [8] において，バス運行時の位置情報が記録されたプローブデータを調査することにより，バスの運行所要時間（移動時間）を予測する取り組みを行ってきた．著者らは他の研究者らによる先行研究や関連研究と対比した特色を 2 点挙げた．

まず，著者らは機械学習の予測精度を上げるために，入力する特徴量として何が最適であるかを探求し，予測に必須の 2 項目（バスの運行区間，およびその区間の実際の運行所要時間）に加えて，予測対象区間の 1 つ前の区間における運行所要時間，および予測対象区間に到達するまでの停車回数を用いると，予測精度が改善するとの結論を得た．また，予測に用いるデータサイズの大小比較についても検証した．他の研究者らによる先行研究や関連研究においては，予測モデルの構築および検証に用いられているデータセットは比較的小規模のものが多かった．著者らの先行研究では，データ数を大きくすることにより，予測精度が改善するかどうかを検証した．各バス停区間ごとにデータ数と予測精度の相関を調査したところ，著者らの予想に反して，相関はほとんどなかった．

先行研究を踏まえて，本研究では，大規模な社会的事象，新型コロナウイルス感染症（COVID-19）が発生したことによる交通環境の変化に伴う，バスの運行所要時間（移動時間）への影響，および，機械学習の予測精度に対する影響に着目した．

本研究を進めるにあたり，2019 年と 2020 年を対比すると，著者らは 4 点の仮説が成り立つと考えた．仮説を Research Question として以下に定義する．

1 九州大学大学院システム情報科学府
福岡市西区元岡 744
2 九州大学大学院システム情報科学研究院
福岡市西区元岡 744
a) kawatani.takuya@m.ait.kyushu-u.ac.jp
b) sato.yuta@m.ait.kyushu-u.ac.jp
c) mine@ait.kyushu-u.ac.jp

©2020 Information Processing Society of Japan

RQ1 バスの運行所要時間が短くなっているのではないか．利用客の減少によりバス停での停車，および乗降にかかる時間が短くなるために起きる遅延，また，道路交通が混雑することによって起きる遅延がなくなるのではないか．

RQ2 RQ1 に関連して，便による所要時間差が小さくなり，バス運行所要時間の分散が減少するのではないか．

RQ3 予測精度に変動が生じるのではないか．機械学習モデルを作成した時点と大きく環境が変わると，予測精度が悪くなるのではないか．

RQ4 機械学習に用いる学習データを多くすることにより，RQ3 の仮説を否定できるのではないか．機械学習に関して，データが多くなれば精度が高くなるという理解が正しいかどうかを確かめられるのではないか．

本研究では，2019 年と 2020 年のそれぞれ 4 月〜6 月における路線バスプローブデータを採取し，統計的な差異を比較した．また，何らかの差異が予想される状況で，差異が生じる前の 2019 年のデータを用いて構築したモデルにより，差異が発生している 2020 年のバスの遅延時間に関する予測が可能であるかどうか，実際に機械学習を実施した．

その結果，RQ1 について，各バス停間のバスの運行所要時間は有意に短くなっていることが判明した．各バス停間の運行所要時間の分布，および平均を求め，その差を統計的に検定したところ，有意差が検出された．RQ2 については，分散の値に有意差があるとはいえないという結果になった．各バス停間の運行所要時間について，その分散の値はわずかに大きくなっていたが，統計的には有意な差であるとはいえなかった．また，RQ3 については，2019 年のデータを用いて作成したモデルによって，2020 年のバス運行所要時間の予測をした場合の予測精度は，2020 年のデータを用いて作成したモデルを用いた場合との差が見られなかった．最後に，RQ4 については，2019 年に採取されたデータすべてを用いて機械学習モデルを構築し，2020 年のデータを予測したところ，2020 年 4 月〜5 月中旬のデータをを用いて，2020 年 5 月下旬〜6 月 10 日のデータを予測したときの予測精度よりも向上はしなかった．

以下，2 節で関連研究を示す．3 節では，本研究に用いた，昭和自動車株式会社の路線バスプローブデータについて説明する．4 節では，本研究に用いたデータの算出法，および統計的検定のための手法について説明する．5 節では，データの分布を観察し，統計的検定を実施した結果を説明する．6 節では結果を考察し，最後に 7 節でまとめと今後の課題を述べる．

2. 関連研究

この節では，関連研究をバスの移動時間予測と，予測に用いる機械学習アルゴリズム，バスの遅延発生要因分析についての研究の 3 つの観点から紹介する．

Bai[3] らは，中国・深セン市の市バスの動的な移動時間予測モデルを提案している．提案モデルはサポートベクター回帰（SVR）またはニューラルネットワーク（ANN）とカルマンフィルタベースのアルゴリズムを用いて構築されている．まず過去のバス走行データから構築された SVR または ANN モデルを用いて基礎となるバスの運行時間を予測し，次にカルマンフィルタベースの動的アルゴリズムを用いて最新のバス運行情報と推定ベースラインの走行時間を用いてバス走行時間を補正する．提案したモデルの性能は、中国・深センの実際のデータで検証された．

山口ら [4], [5] は，80 以上のバス路線を扱い，線形回帰（LR），ニューラルネットワーク（ANN），サポートベクター回帰（SVR），ランダムフォレスト（RF），勾配ブースティング決定木（GBDT）の 5 つの機械学習手法を用いて，隣接するバス停区間における，定刻の移動時間と実際の移動時間の差である遅延を予測するモデルを提案している．5 手法の比較の結果，GBDT の優位性を示した．

田中ら [6] は，静岡県富士市のバス路線を対象として，交通状況による遅延を考慮し円滑な乗換を保証するバス時刻表を設計提案している．田中らは，バスの遅延を発生させる主要因として，走行中に発生する渋滞遅延と，バス停における乗客の乗降が原因で発生する乗降遅延があることを確認している．

上記の先行研究に対して，本研究の特色は，COVID-19 発生前と，COVID-19 が発生している現状との間において，実際のバス運行状況の変化が発生したかどうかをプローブデータから調査・比較している点にある．バス路線がとりまく環境が変化したときに発生する事象の一部，特に，運行所要時間の変化について，その性質を統計的に明らかにした．

3. データセット

3.1 プローブデータの特性

この節では，本研究において使用したデータセットについて述べる．本研究では，筆者らの先行研究 [7], [8] と同様に，昭和自動車株式会社（本社・佐賀県唐津市）が所有する路線バス車両のうち，調査対象である福岡地区の九州大学線で運行される 44 台のプローブデータを利用した．九州大学線は，九大学研都市駅と九州大学伊都キャンパスを結ぶ全長約 4 km のバス路線である．当路線はおおまかに分けて 3 つのルートから成る．路線図を**図 1** に示す．

このプローブデータには，車両 ID，走行時刻，緯度，経度，速度が記録されている．詳細を**表 1** に示す．

ここで，データの特性を述べる．記録されている速度について，その最小値は 9 km/h である．つまり，完全に停止している状態であるかどうかは，データからは読み取れない．さらに，データの記録間隔が一定でない．原則として，走行中は一定の距離間隔でデータが記録されるが，進

©2020 Information Processing Society of Japan

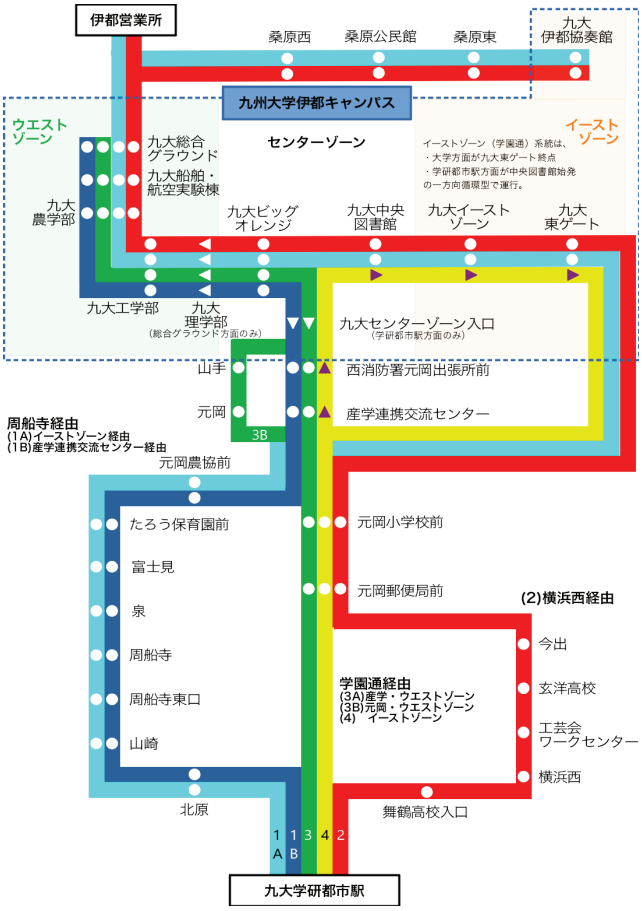

図 1　昭和自動車九州大学線路線図

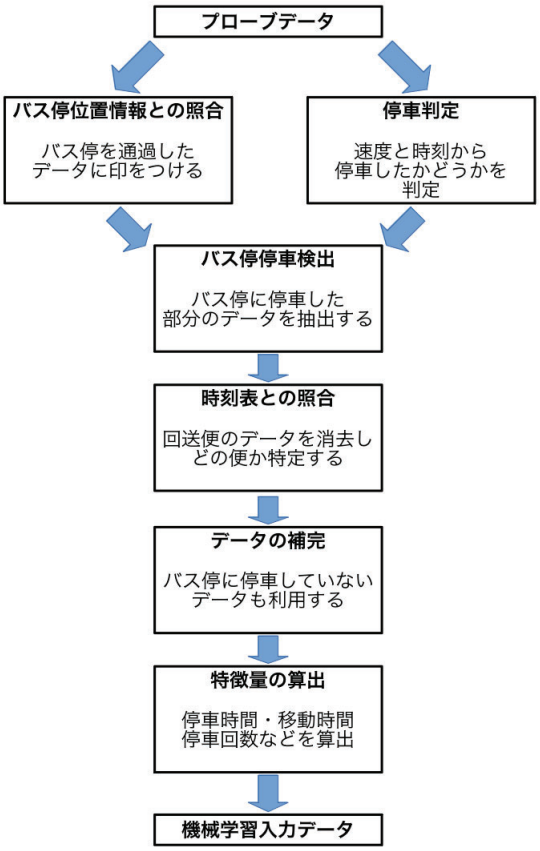

図 2　特徴量算出手順の概略

表 1　走行記録データの書式

項目	概要	例
ID	車両を識別する数字	11
走行時刻	位置情報が記録された日時	2019/05/21 11:58:15
緯度	度数法・小数点なし	3360033
経度	度数法・小数点なし	13021238
速度	車両の速度（km/h）	37

表 2　バス運行本数の変化

日時	平日	土曜	日祝
2019.04.01 改正	444	204	174
2019.10.01 改正	366	167	157
2020.04.01 改正	361	162	152
2020.04.13 特別減便	331	162	152
2020.05.07 特別減便 II	134	134	134
2020.05.21 変更	314	144	134

行方向や加速度が大きく変化したときは，より細かい間隔で記録される．

本研究に用いたデータの採取期間は，2019 年 4 月 1 日〜12 月 31 日，また，2020 年 4 月 1 日〜6 月 10 日である．当該期間におけるバス運行本数の変化を**表 2** に示す．2020 年 4 月 13 日から 5 月 20 日までの間，所定の 2020 年 4 月 1 日改正運行ダイヤから運行本数が臨時に削減されていた．5 月 7 日から 5 月 20 日まで，最も本数が削減されていた時期は，所定の運行ダイヤから平日の場合でおよそ 63%，土曜・日祝日においておよそ 17%分本数が減少していた．

3.2　特徴量の算出

プローブデータから，機械学習に用いるために，著者らの先行研究 [8] と同様に特徴量を算出，および追加した．

特徴量の算出手順の概略を示す．全体像は**図 2** の通りである．

特徴量の算出手順は次の通りである．

(1) バス停に停車したことを検出する．
(2) プローブデータシーケンスがどの運行便か特定する．
(3) バス停を通過したデータを用いて各バス停間のデータを補完する．
(4) 停車時間・移動時間・停車回数などを算出する．

バス停に停車したことを検出する方法を述べる．方法は筆者らの先行研究 [8] と同様である．

3.1 節で述べたように，プローブデータの速度情報は，9 km/h 未満が記録されない．そのため，バスがバス停で停車したかどうかを判定するために，次の方法でバス停のそばを通ったかどうかを検出した．また，プローブデータから停車したと思われる区間を検出し，停車時間を予測した．

まず，あるバス停の 100 m 四方の範囲内で記録されているデータを見つける．次に，プローブデータのシーケンス

©2020 Information Processing Society of Japan

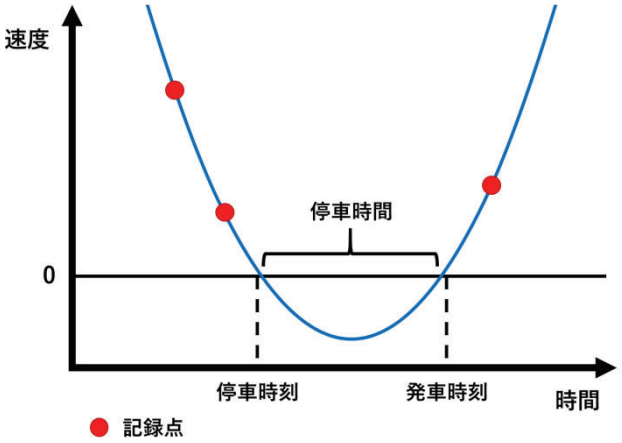

図 3 停車時刻・発車時刻・停車時間の計算

からバスの進行方向を求める．バスの進行方向とバス停が立つ方向とが一致したときに，そのバス停を通ったと記録する．なお，バス停に停車したかどうかを判定するため，30 km/h 以下の速度で通過した場合のみを検出するようにした．停車時間の予測には，**図 3** に示す時間と速度の 2 次関数を考えた．連続する 3 つの記録点において，その時間と速度を通る 2 次関数を求め，停車時刻，発車時刻，停車時間を算出した．

次に，どの運行便であるかを特定し，回送便をデータから除去するために時刻表との照合を行った．これにより，ある運行データが定刻で何時何分の便であるのかを求めることができる．

第三に，各バス停間のデータを補完した．バスはすべてのバス停に停車することは少なく，いくつかのバス停を通過することは一般的である．たとえば，バス停 A を出発してバス停 B, C を通過してバス停 D に停車したとする．本研究においては隣接バス停間のデータを用いて運行所要時間（移動時間）を予測するため，そのままでは「バス停 A からバス停 D までの運行データ」は用いることができない．ここで，バス停間の距離の比で各バス停間の移動時間を分割配分することにより，「バス停 A, B 間の運行データ」「バス停 B, C 間の運行データ」「バス停 C, D 間の運行データ」として用いた．

最後に，各バス停区間のデータが得られたことから，そのバス停区間の 1 区間前のバス停間の移動時間，当該バス停区間に到達するまでに当該便が何回停車したか，また，停車時間の合計を計算して追加した．

算出した特徴量を次に示す．

- どのバス停間のデータかを示す，バス停 ID の組．
- バスの運行曜日．
- 予測対象区間における，出発バス停の定刻出発時刻．
- 予測対象区間の，1 区間前の運行所要時間．
- 予測対象区間に到達するまでの停車回数．
- 予測対象区間に到達するまでの総停車時間．

表 3 学習用データとテストデータの分割

区分	学習データの期間	テストデータの期間
A	2019.04.01 - 2019.05.20	2019.05.21 - 2019.06.10
B	2020.04.01 - 2020.05.20	2020.05.21 - 2020.06.10
C	2019.04.01 - 2019.05.20	2020.05.21 - 2020.06.10
D	2019.04.01 - 2019.06.10	2020.04.01 - 2020.06.10
E	2019.04.01 - 2019.12.31	2020.04.01 - 2020.06.10

4. 実験

まず，RQ1 および RQ2 を検証するため，プローブデータを統計的に比較した．プローブデータについて，採取されたバスの台数，データ数，バス停間の運行所要時間（移動時間）の平均，中央値，および分散を算出した．移動時間の分布に有意差が存在するかどうかを統計的に検定するため，ヒストグラムを用いて分布の概形を把握し，また，クラスカル・ウォリス順位和検定 [9], [12] を用いた．ここで，クラスカル・ウォリス順位和検定とは，2 群以上の分布に差があるかどうかを調べる，ノンパラメトリックな検定手法である．さらに，移動時間の分散に差があるかどうかを調べるために，ルビーン検定 [10] を用いた．ここで，ルビーン検定とは，2 群以上の分散に差があるかどうかを検定する手法である．

次に，RQ3 と RQ4 を検証するため，機械学習手法を適用して，予測精度の変化を観察した．機械学習モデルを構築するにあたり，3 節で示した特徴量のうち，「予測対象区間の 1 区間前の運行所要時間」と「予測対象区間に到達するまでの停車回数」を採用してモデルを構築した．この 2 つの特徴量は，著者らの先行研究 [8] において，予測精度をもっとも向上させる効果があると確認できたものである．

学習データとして用いる期間と，テストデータとして用いる期間を変えて学習およびテストを行った．機械学習手法としては，山口の先行研究 [4], [5] で最適だと判定された，GBDT（勾配ブースティング決定木）手法を適用した．GBDT アルゴリズムについては，ハイパーパラメータのチューニングをしない状態でも比較的よい予測精度を示す，CatBoost を選択した．CatBoost のハイパーパラメータはチューニングをしていない，デフォルトの状態で実験した．

学習用データとテストデータの分割は，**表 3** に示す 5 通りを採用した．この 5 通りを採用した理由を以下に示す．

4.1 A 対 B: 各年における予測精度の比較

A では，2019 年の 4 月〜6 月のデータを学習データとテストデータに分割して機械学習および評価を実施した．また，B では 2020 年の 4 月〜6 月のデータについて，A と同様に機械学習と評価を実施した．これにより，まず年度の違いによる予測精度の違いを見ることができると考えた．

©2020 Information Processing Society of Japan

4.2 A 対 B 対 C: 前年の機械学習モデルによる，今年の予測精度の算出

C は，A の学習データを用いた機械学習モデルにより，B のテストデータの予測をしている．これは，前年度のデータを用いて構築した機械学習モデルと，当年度のデータを用いて構築した機械学習モデルと，どちらが当年度の予測に有用であるかどうかを試すものである．

4.3 C 対 D: 期間を同一として，前年のモデルにより今年を予測した場合

D は，前年度のデータを用いて構築したモデルの予測精度を比較する際に，学習データ数を増加させた場合の比較となる．

4.4 D 対 E: 学習データの期間を長くした場合の比較

E は，D と比較してさらに学習データの期間を長くとって機械学習モデルの構築を行った場合を検討する．E は単にデータ数が多いのみならず，春から冬にかけての季節的な変動要因も学習データ内に含まれていることが推測される．これにより，路線をとりまく環境をより反映した機械学習モデルが構築されるため，より良い予測精度となることが期待できる．

予測精度の評価指標には，MAE（平均絶対誤差），および MAAPE（Mean Arctangent Absolute Percentage Error）[11] を利用した．MAE は路線バスの運行所要時間を予測する先行研究において，予測精度を検証する指標として広く用いられている．また，MAAPE は MAPE（平均絶対パーセント誤差）を改良した指標であり，これもまた予測精度を検証するのに有用である．本研究では，MAAPE を主に用いた．以下に，MAE と MAAPE の定義を数式で示す．ここで，n はデータ数，y_i は実際のバス停間移動時間，$\hat{y}_i$ は機械学習によって予測されたバス停間移動時間である．

$$MAE = \frac{1}{n}\sum_{i=0}^{n-1}|y_i - \hat{y}_i|$$

$$MAAPE = \frac{100}{n}\sum_{i=0}^{n-1}\arctan\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right)$$

5. 結果

5.1 統計的な比較

2019 年と 2020 年の 4 月 1 日〜6 月 10 日における，データ数と基本統計量を**表 4** 〜 **表 6** に示す．プローブデータから得られた，各年度の 4 月から 6 月におけるバスの走行台数は，ほぼ変化がなかった．データ数を比較すると，2019 年と比べて，2020 年はデータ数が半減していた．区間ごとのデータ数を観察したところ，多くの区間においても全体データ数と同様に，データ数がおよそ半減している結果となった．

表 4 基本統計量

項目	2019 年	2020 年
走行台数	41 台	39 台
データ数	65,323	33,407
移動時間の平均 [s]	97.840	94.596
移動時間の分散 [s^2]	7092.383	7203.912
移動時間の中央値 [s]	73.915	69.564
全バス停区間で算出した MAE [s]	15.045	13.997
全バス停区間で算出した MAAPE	14.4%	12.9%

表 5 九大学研都市駅に到着するまでの停車回数

項目	2019 年	2020 年
合計	400	21
平均	0.276	0.036
分散	1.908	0.284
中央値	0	0
最大値	17	8
最小値	0	0

表 6 九大学研都市駅に到着するまでの総停車時間

項目	2019 年	2020 年
平均	7.409	0.929
分散	1971.691	106.739
中央値	0	0
最大値	834.35	189.04
最小値	0	0

次に，バスの移動時間について，平均と分散，中央値を求めた．ここでいう移動時間の平均とは，各バス停区間の所要時間をすべて足し合わせて，データ数×バス停区間数の数値で割った値である．

バスが予測対象区間に到達するまでの停車回数についての統計量を調べた．ここでは，バスが九大学研都市駅に到着するまでの停車回数について求めた．**表 5** に示すように，2020 年は，バス停車回数の合計が大きく減少した．平均停車回数や分散も減少していた．また，バスが九大学研都市駅に到着するまでの総停車時間を見ると，表 6 に示すとおり，2019 年と比べて 2020 年は減少していた．

表 4 のとおり，移動時間の平均は 97.840 s から 94.596 s と値が小さくなっていた．また，中央値は 73.915 s から 69.564 s と，こちらも値が小さくなっていた．中央値の変動に統計的な有意差が存在するかどうか確かめるために，まず，ヒストグラムを作成して概形を観察した．ヒストグラムを**図 4** と**図 5** に示す．このヒストグラムからわかるとおり，バス停間の移動時間平均は，左に大きく偏った分布となった．また，移動時間 300 s の付近に小さなピークが現れていた．

ヒストグラムの概形から，移動時間の分布に正規性を仮定できないため，2019 年と 2020 年の間に分布の差があるかどうか，クラスカル・ウォリス順位和検定により検定し

©2020 Information Processing Society of Japan

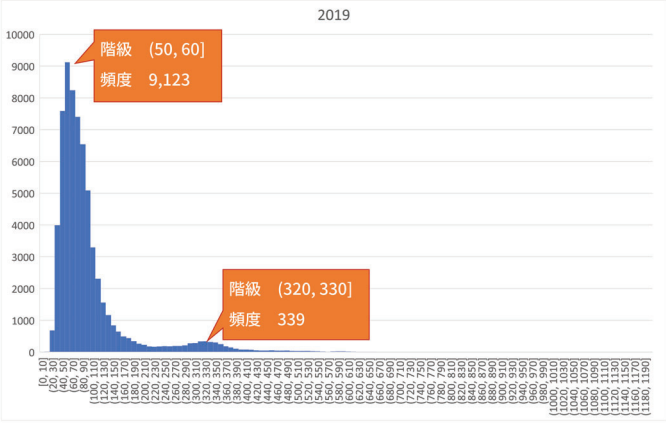

図 4 移動時間分布（2019 年）

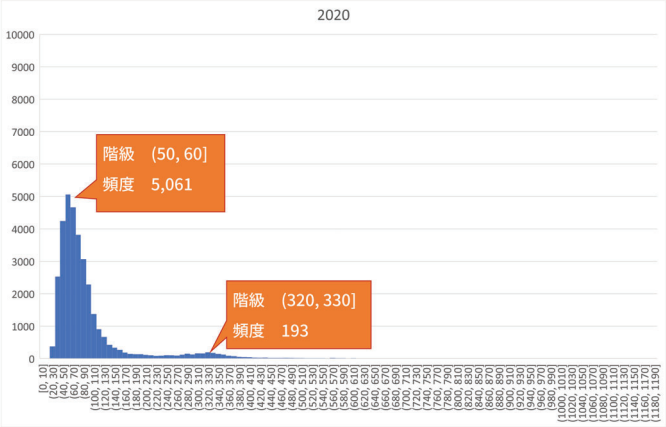

図 5 移動時間分布（2020 年）

た．その結果，p 値は 4.89×10^{-55} となり，移動時間の分布に有意差がみられる結果となった．中央値の変動から，移動時間は全体的に短くなったと考えられる．

移動時間の分散については，表 4 に示すとおり，2020 年の方がやや数値が大きくなっていた．この分散の値の差が有意であるかどうか，ルビーン検定により検定した．その結果，p 値は 0.073 となり，有意水準 5%で有意差があるとはいえない結果となった．この分散の値から標準偏差を算出して考察すると，おおむね各バス停間の所要時間の大小に一致する結果となっていた．

5.2 機械学習

次に，機械学習を行い，予測精度を算出した結果を示す．

機械学習にあたり，先行研究 [8] の成果を参照し，予測に次の 5 点の特徴量を用いた．

- バスがどのバス停間を運行しているか．
- バスが予測対象区間の起点バス停を出発する定刻の時刻．
- 当該バス停間における，実際の運行所要時間．
- 予測対象区間の 1 つ前の区間における運行所要時間
- 予測対象区間に到達するまでの停車回数

5.2.1 A 対 B: 各年における予測精度の比較

表 8 における区分 A と区分 B を比較した．全区間を対象に算出した MAE は，2019 年が 15.045 s，2020 年が 13.997 s となった．これは，2019 年に比べて，2020 年の予測誤差が小さくなったことを示している．また，MAAPE は，2019 年が 14.4%，2020 年が 12.9%で，2020 年の方が値が小さくなっていた．MAAPE の値が小さいのは，予測精度が良いことを意味する．バス停区間ごとの MAAPE を比較すると，全 78 区間中 49 区間で，2019 年よりも 2020 年の方が予測精度が良い結果になった．

5.2.2 A 対 B 対 C: 前年の機械学習モデルによる，今年の予測精度の算出

区分 A と区分 B に対して，区分 C を比較した．全区間を対象に算出した MAAPE は，A よりも精度が良く，B よりは精度が悪い結果になった．また，各バス停区間の MAAPE も同様に，A よりも精度が良く，B よりは精度が悪かった．

5.2.3 C 対 D: 期間を同一として，前年のモデルにより今年を予測した場合

区分 C と区分 D を比較した．データ数が増えれば単純に精度が改善するという仮説は成り立たず，他の 4 つの試行と比較しても，最も MAAPE の値が大きくなった．

5.2.4 D 対 E: 学習データの期間を長くした場合の比較

最後に，区分 D と区分 E を比較した．両者の差は，学習データに用いたデータの期間の長さである．全 78 区間中 54 区間で MAAPE の値が良くなっていた．

6. 考察

6.1 統計的分析結果についての考察

統計的分析を行った結果について考察する．2019 年と比較して，2020 年のデータ数が半減していた理由は，バスの運行本数が通常ダイヤにおいても減少し，さらに COVID-19 の影響により特別減便が実施されていたことが，データにも現れたものであると考えられる．各バス停区間で見た場合も，データ数が半減していることから，特定の区間だけで運行本数が減少したのではなく，全体的に減少したことがわかる．

バスが予測対象区間に到達するまでの停車回数や総停車時間が大幅に減少した理由として，2 つの可能性が考えられる．ひとつは，バスの利用客減少によりバス停への停車回数が減少したことである．もうひとつは，交通混雑の減少により，バス停以外での停車も減少したことである．

移動時間の分布が左に大きく偏り，そのピークが 60 s に

©2020 Information Processing Society of Japan

表 7 表 3 で示した異なるデータを用いた場合の，区間ごとの MAE の比較

区間	区分 A	区分 B	区分 C	区分 D	区分 E
（全区間）	15.045 s	13.997 s	14.419 s	14.132 s	13.866 s
九大学研都市駅→北原	29.088 s	11.652 s	14.023 s	16.425 s	14.823 s
九大理学部→九大工学部	7.398 s	6.791 s	8.631 s	9.031 s	8.370 s
富士見→泉	7.678 s	6.680 s	8.026 s	8.347 s	7.867 s
西消防署→産学連携交流センター	7.294 s	7.427 s	7.863 s	7.506 s	7.914 s

表 8 表 3 で示した異なるデータを用いた場合の，区間ごとの MAAPE の比較

区間	区分 A	区分 B	区分 C	区分 D	区分 E
（全区間）	14.3%	12.9%	14.6%	14.9%	14.4%
九大学研都市駅→北原	36.4%	20.7%	24.1%	24.4%	22.0%
九大理学部→九大工学部	14.7%	14.8%	17.1%	17.6%	16.8%
富士見→泉	11.8%	10.5%	11.8%	12.3%	11.7%
西消防署→産学連携交流センター	15.8%	16.7%	17.3%	16.5%	17.4%

存在していて，大部分のデータが 180 s 以下となっていた．これは，調査対象である九州大学線のバス停の間隔が，おおむね所要時間 1 分で，大部分が 3 分以内であるという，時刻表から得られた結果に合致する．また，移動時間 300 s 付近に小さなピークが存在している点は，バス停の間隔が所要時間 5 分程度だけ離れている箇所が 1 箇所存在することを表していると考えられる．

2019 年と 2020 年のデータを比較して移動時間の分布に有意差が見られ，2020 年のデータの中央値が小さくなっていたことから，2020 年は 2019 年と比べて，バス停間の移動時間が全体的に短くなったと考えられる．

6.2 機械学習の結果についての考察

5.2 節で，学習データとテストデータの期間を変えつつ，予測精度を算出した．ここで，MAAPE の値の大小をどのように評価するか考察する．基本的に，MAAPE の値が小さいほど予測精度が良いと評価できる．当初，著者らの考えでは，機械学習の学習データを多くするほど，予測精度が高まり，MAAPE の値が小さくなると推測していた．しかし，実際にはその推測とは異なり，予測対象の期間に近い，短い期間のデータを使って学習したモデルが，MAAPE の値が最も小さい結果となった．

この結果について，2 つの仮説を検討する．ひとつは，「予測したい時期の直近の時期にあたる過去データは，予測したい時期の交通の傾向や気象条件などの要因が似通っているため，予測精度が高まる」という仮説である．これは直近ほど予測しやすく，離れた時期になるほど予測が難しくなるという直感的な認識に合っている．しかし，この仮説では区分 A と区分 B の差，すなわち 2019 年のデータで機械学習を行い，同じく 2019 年を予測した MAAPE と，2020 年のデータで学習し，2020 年を予測した MAAPE に差が生じている理由が説明できない．

ここで，もうひとつの仮説「外的要因の変化により，予測精度が変動した」説を考える．文献 [1], [2] が示すように，2020 年 4 月〜6 月の期間中は，バスの利用客数が大幅に減少し，道路交通量も減少していた．調査対象の九州大学線においても，大学の講義がオンライン講義に切り替わり，大学へ行き来する人の数が大幅に減少していたことから，バス利用客数も大きく減少していると推察される．バスが遅延する要因として考えられるものには，道路混雑状況のほかに，バス停における利用客の乗降がある [6]．道路の混雑が緩和され，かつ，利用客の減少によりバス停への停車時間が減少したことでバスの移動時間予測に関する不確定要素が減り，結果として予測精度が改善した可能性が考えられる．

これらを踏まえて，1 節で設定した Research Question に対する結論を述べる．

RQ1 に対する結論 2019 年と 2020 年の同時期を比較すると，バスの運行を示すデータ数は半減していた．また，バス停間移動時間は，2020 年の方が有意に短くなったと考えられる．

RQ2 に対する結論 移動時間の分散には 5 %水準で有意差が見られなかった．

RQ3 に対する結論 2019 年の学習データを用いて構築した機械学習モデルにより，2019 年のテストデータと 2020 年のテストデータに対する予測精度を比較すると，2020 年のテストデータに対する精度が MAAPE の値において 0.3 ポイント悪くなっていた．つまり，機械学習モデル作成時と予測対象時期の環境が変わると，精度が低下する可能性がある．

RQ4 に対する結論 学習データの量（期間）を多くとり，同じテストデータに対する予測精度を比較した結果，学習データの量が多いほうが MAAPE の値が小さくなっていた．

7. まとめと今後の課題

本研究では，新型コロナウイルス感染症（COVID-19）の影響により，路線バスの利用客数や道路交通量が減少した

©2020 Information Processing Society of Japan

2020 年 4 月〜6 月の状況が，路線バスの運行所要時間（移動時間）に与える影響を調査した．また，機械学習による移動時間予測の精度がどう変動するかを観察した．

2020 年 4 月〜6 月の運行状況を，2019 年の同時期と比較して統計的に分析したところ，運行本数の減少がデータ数の減少という形でも現れていた．また，バス停間の移動時間が有意に短くなっていた．この理由として，バスの遅延の原因となるバス停での停車回数および停車時間が減少したこと，また，交通混雑の減少によりバス停以外での停車回数も減少したことが考えられる．

予測精度の比較実験において，2019 年のデータを用いた機械学習および予測の実験と，2020 年のデータを用いた実験との比較（区分 A と区分 B の比較）では，2020 年のデータを用いた実験（区分 B）のほうが MAAPE が小さい結果となった．これは，道路の混雑が緩和され，かつ，利用客の減少によりバス停への停車時間が減少したために，移動時間予測に関する不確定要素が減り，結果として予測精度が改善した可能性が考えられる．

最後に，本研究では，予測精度の向上に寄与する特徴量や，バス停区間ごとに予測精度の差が生じる要因が未解明である．また，予測精度が実用上十分であるかどうかを検討する必要がある．これらは今後の課題としたい．

謝辞

本研究の一部は，JP20H01728 の支援を受けた．本研究にあたり，昭和自動車株式会社には，バスの走行記録データ取得の実証実験に協力を頂き，データ分析の許可を頂いた．感謝申し上げる．

参考文献

[1] 国土交通省九州運輸局：新型コロナウイルス感染症による九州の公共交通事業者への影響と対策事例 (入手先 ⟨https://wwwtb.mlit.go.jp/kyushu/press/00001_00339.html⟩) (2020.07.29).

[2] 国土交通省報道発表資料：ETC2.0 プローブデータによる交通状況分析について (入手先 ⟨https://www.mlit.go.jp/report/press/road01_hh_001323.html⟩) (2020.07.29).

[3] C. Bai, Z. Peng, Q. Lu, J. Sun, *Dynamic Bus Travel Time Prediction Models on Road with Multiple Bus Routes*, Computational Intelligence and Neuroscience Vol. 2015, 432389 (2015).

[4] T. Yamaguchi, M. As, and T. Mine, *Prediction of bus delay over intervals on a various kinds of routes using bus probe data*, in The 5th IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT 2018), pp. 97–106 (2018).

[5] T. Yamaguchi, R. Maita, T. Kawatani, T. Mine, *Prediction of travel time over intervals between two bus stops using bus probe data*, Proceedings of 17th ITS Sympodium 2019, (2019).

[6] 田中 健裕, 高松 瑞代, 菅原 宏明, 田口 東,「交通状況による遅延を考慮し円滑な乗換を保証するバス時刻表の設計」, 都市計画論文集 vol. 52, No. 3, pp. 1320–1326, (2017).

[7] 川谷 卓哉, 舞田 涼太朗, 山口 翼, 峯 恒憲：バスプローブデータからの運行特性発見 (情報処理学会研究報告高度交通システムとスマートコミュニティ（ITS）,2019-ITS-79(17),1-6 (2019-11-13) , 2188-8965.

[8] T. Kawatani, T. Yamaguchi, Y. Sato, R. Maita, T. Mine, *Prediction of bus travel time over intervals between pairs of adjacent bus stops using city bus probe data*, 第 18 回 ITS シンポジウム 2020, (2020).

[9] Kruskal; Wallis: *Use of ranks in one-criterion variance analysis*, Journal of the American Statistical Association (1952).

[10] Levene, Howard.: *Robust tests for equality of variances*, Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (1960).

[11] S. Kim, H. Kim, *A new metric of absolute percentage error for intermittent demand forecasts*, International Journal of Forecasting, Vol. 32, Issue 3, pp. 669–679, (2016).

[12] 石村貞夫，石村光資郎：入門はじめての分散分析と多重比較，東京図書 (2008).

©2020 Information Processing Society of Japan