

路線バス等の運行状況改善に向けた位置情報履歴 データ解析に関する研究

川谷, 卓哉

<https://hdl.handle.net/2324/6787633>

出版情報 : 九州大学, 2022, 博士 (情報科学), 課程博士
バージョン :
権利関係 :

路線バス等の運行状況改善に向けた位置情報履歴
データ解析に関する研究

川谷 卓哉

概要

本研究は、社会インフラのひとつとして不可欠な公共交通機関である路線バスについて、車両の位置情報を大量に収集蓄積した位置情報履歴データを解析し、また、機械学習を利用することで、バスの運行所要時間予測を低コストかつ高い精度で実現する手法を示した。また、路線バス等の自動車の安全運行実現につながる要素のひとつとして急ブレーキに着目し、その発生要因を位置情報履歴データと機械学習を組み合わせて解明する手法を提示した。

本研究では第一に、バスの運行所要時間を予測するために必要な情報や、その収集手法を明らかにした。第3章では、約1年間にわたる路線バス車両の位置情報履歴データを用いて、隣接するバス停間1区間の運行所要時間を予測する機械学習モデルを構築し、高い予測精度を実現する特徴量として「予測対象区間の1つ前のバス停間の運行所要時間」と「予測対象区間に到達するまでのバスの停車回数」が有用であることを解明した。しかし、これらの特徴量はリアルタイム性の高い情報であり、本研究で利用している位置情報履歴データ収集システムでは実際に取得することが難しく、また、その情報を解析する計算時間の観点からも利用することは困難であった。そのため、本研究のシステムで取得し利用可能なよりリアルタイム性の低い情報として「予測対象区間における1時間前の運行所要時間実績値の平均値」に着目した。この情報を利用することで、リアルタイムな情報が利用できない状況においても、リアルタイムな情報が必要であると示す、バス運行所要時間予測に関する多くの先行研究と比較して遜色のない予測精度を達成した。

また、第4章では、路線バスの位置情報履歴データを収集する設備について、より高い予測精度を実現するために収集用路側機を増設する場合の設置箇所の選定手法を示した。まず、リアルタイム性の低い情報を利用して予測する場合に有効だった特徴量「予測対象区間における1時間前の運行所要時間実績値の平均値」が取得できていない状況が一部の時間帯において発生していること、また、この状況を改善するためには、路側機を路線のもう一方の端点へ設置すればよいと明らかにした。さらに、リアルタイムにより近い情報を追加で取得利用して予測精度を向上させる際の設置箇所選定基準として、運行所要時間予測にあたっての特徴量重要度の値が大きいバス停間の運行所要時間、その中でも起点側

に近い区間の運行所要時間を得られるような地点に設置することが望ましいとした。

第 5 章では、会津若松市公用車・公共交通車両走行情報オープンデータセットに対して、サポートベクターマシンを利用した機械学習と、特徴選択を組み合わせた手法により、自動車の交通安全に関連する事象のひとつである「急ブレーキ」を識別する機械学習モデルを生成し、急ブレーキ事象の識別に寄与した特徴量を解析することで急ブレーキの発生条件の抽出を行った。識別に寄与した特徴量の組を調査した結果、発生地点に関する特徴量よりも、日時や曜日などの時間的要因に関する特徴量が有効であると明らかにした。急ブレーキの識別に有効であった上位の特徴量の特徴量ラベルが指し示す条件を満たす位置を地図上にプロットした結果、急ブレーキ事象のデータ件数をもとにして得られた位置とは異なることが判明した。

本研究では、運行所要時間予測に必要な位置情報履歴データがリアルタイムに得られず、1 時間ごとのように間欠的な収集しかできない場合でも問題なく、リアルタイムな情報が得られる場合と遜色のない予測手法を開発できた。さらに、予測にあたっては、位置情報履歴データの他は、バス停の位置情報を別途組み合わせるだけでよいような手法を構築することができた。このことは、バスの詳細な運行状況を記録する機器がなくとも、汎用の位置情報計測機器、およびデータ収集機器から成るシステムがあれば問題ないことを示している。その際のコストは低廉であり、日本国内のみならず世界で広く適用可能である。

本研究では、運行所要時間予測にあたって、予測に有用なリアルタイム性の低い情報として採用した「予測対象区間における 1 時間前の運行所要時間実績値の平均値」の選定基準として、「1 時間前」としたことが妥当であるかの検証が十分でなく、本研究の対象路線よりもネットワークの複雑度が増したバス路線へ適用するにあたって必要な追加条件の検討ができていないため、今後の課題として検証したい。

目次

第 1 章	序論	1
1.1	研究背景	1
1.2	研究目的	4
1.3	本研究の独自性と貢献	4
1.4	本論文の構成	5
第 2 章	導入	7
2.1	はじめに	7
2.2	位置情報履歴データ収集	7
2.3	位置情報履歴データ変換, 統計量および特徴量算出	9
2.4	データセット 1: 会津若松市公用車・公共交通車両走行情報	9
2.5	データセット 2: 昭和バス伊都営業所管内路線バス位置情報履歴データ	10
2.6	予測精度評価指標	14
第 3 章	バス運行所要時間の予測	19
3.1	はじめに	19
3.2	関連研究	21
3.3	データセット	23
3.4	実験手法	24
3.5	結果	25
3.6	理想的な特徴量が得られない場合	28
3.7	先行研究との予測精度の比較	31
3.8	考察	32
3.9	まとめ	34
第 4 章	低コストと高予測精度を実現するデータ追加収集箇所選定手法	37
4.1	はじめに	37

4.2	関連研究	38
4.3	問題の定義	39
4.4	データセット	40
4.5	解析手法	40
4.6	解析 I: 所要時間予測精度を向上させる区間と当該区間の所要時間の分散 の関係	42
4.7	解析 II: 1 時間前の所要時間実績値を取得できる地点の選定	48
4.8	考察	55
4.9	まとめ	56
第 5 章	急ブレーキ事象の識別と要因解明	59
5.1	はじめに	59
5.2	データセット	60
5.3	危険な走行挙動「急ブレーキ」	60
5.4	急ブレーキ発生状況の基礎解析	60
5.5	機械学習手法を用いた急ブレーキの識別	62
5.6	実験	66
5.7	まとめ	68
第 6 章	まとめ・今後の課題	75
6.1	本研究の貢献	75
6.2	本研究における限界	77
6.3	今後の課題	78
	参考文献	81

目次

2.1	昭和バス九州大学線路線図（2019年4月1日現在）	12
3.1	特徴量の概念図	24
3.2	特徴量 A, B を追加しない場合の MAAPE	27
3.3	特徴量 A, B を追加した場合の MAAPE	27
4.1	1 時間以内に当該時間帯の運行所要時間情報が得られなかった件数（学園通経由・九州大学→九大学研都市駅）	54
5.1	急ブレーキ発生件数の上位 50 地点	62
5.2	急ブレーキの発生確率と発生件数の関係	67
5.3	特徴選択を行った場合の急ブレーキ識別性能（地点の特徴量を除いたとき）	67
5.4	特徴選択を行った場合の急ブレーキ識別性能（地点の特徴量を除いたとき）	72
5.5	急ブレーキ正例の特徴語 “kym:10:2017:02” にあてはまる地点上位 50 件	74

表目次

2.1	「会津若松市公用車・公共交通車両走行情報」データ諸元	10
2.2	「会津若松市公用車・公共交通車両走行情報」データ書式	10
2.3	「昭和バス伊都営業所管内路線バス位置情報履歴データ」データ諸元 . . .	11
2.4	「昭和バス伊都営業所管内路線バス位置情報履歴データ」データ書式 . . .	11
2.5	九州大学線の運行本数の変化	11
2.6	GRS80 楕円体の長半径と扁平率	14
2.7	ルート別シーケンス数（九州大学線）	14
2.8	混同行列	15
3.1	データの概要	23
3.2	予測精度評価	26
3.3	特徴量と MAAPE との相関係数	27
3.4	予測精度比較（MAE・単位は秒）	29
3.5	予測精度比較（MAAPE）	29
3.6	特徴量重要度（パターン A・学園通経由・区間所要時間を加えない） . . .	30
3.7	特徴量重要度（パターン A・学園通経由・大学出口まで加える）	30
3.8	特徴量重要度（パターン B・学園通経由・区間所要時間を加えない） . . .	30
3.9	特徴量重要度（パターン A・周船寺経由・区間所要時間を加えない） . . .	30
3.10	特徴量重要度（パターン A・周船寺経由・大学出口まで加える）	31
3.11	特徴量重要度（パターン B・周船寺経由・区間所要時間を加えない） . . .	31
3.12	特徴量重要度（パターン A・横浜西経由・区間所要時間を加えない） . . .	31
3.13	特徴量重要度（パターン A・横浜西経由・大学出口まで加える）	31
3.14	特徴量重要度（パターン B・横浜西経由・区間所要時間を加えない） . . .	32
3.15	先行研究の予測精度との比較（MAPE）	32
4.1	データセット概要	40
4.2	解析に利用した特徴量	41

4.3	運行所要時間の分散が大きい区間（九州大学方面）	42
4.4	運行所要時間の分散が大きい区間（九大学研都市駅方面）	42
4.5	指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（学園通經由九州大学方面）	43
4.6	指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（周船寺經由九州大学方面）	43
4.7	指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（横浜西經由九州大学方面）	44
4.8	指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（学園通經由九大学研都市駅方面）	44
4.9	指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（周船寺經由九大学研都市駅方面）	45
4.10	指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（横浜西經由九大学研都市駅方面）	46
4.11	特微量重要度（学園通經由九州大学方面）	46
4.12	特微量重要度（周船寺經由九州大学方面）	47
4.13	特微量重要度（横浜西經由九州大学方面）	47
4.14	特微量重要度（学園通經由九大学研都市駅方面）	47
4.15	特微量重要度（周船寺經由九大学研都市駅方面）	47
4.16	特微量重要度（横浜西經由九大学研都市駅方面）	48
4.17	指定したバス停までの各区間所要時間を加えた場合の予測精度（学園通經由九州大学方面）	48
4.18	指定したバス停までの各区間所要時間を加えた場合の予測精度（周船寺經由九州大学方面）	49
4.19	指定したバス停までの各区間所要時間を加えた場合の予測精度（横浜西經由九州大学方面）	49
4.20	指定したバス停までの各区間所要時間を加えた場合の予測精度（学園通經由九大学研都市駅方面）	50
4.21	指定したバス停までの各区間所要時間を加えた場合の予測精度（周船寺經由九大学研都市駅方面）	51
4.22	指定したバス停までの各区間所要時間を加えた場合の予測精度（横浜西經由九大学研都市駅方面）	52
4.23	2019 年度運行所要時間取得状況（九大学研都市駅→九州大学）	52
4.24	2019 年度運行所要時間取得状況（九州大学→九大学研都市駅）	53

4.25	2019 年度運行所要時間統計値（九大学研都市駅→九州大学・異常値除去後）（秒）	53
4.26	2019 年度運行所要時間統計値（九州大学→九大学研都市駅・異常値除去後）（秒）	53
4.27	1 時間前の運行所要時間を取得できた数（学園通経由・九州大学→九大学研都市駅）	54
5.1	急ブレーキ発生地点と急ブレーキの割合	61
5.2	SVM+FS モデル構築に利用したデータ数	63
5.3	選択した特徴量	63
5.4	急ブレーキに影響する地点の要因の効果（学習用データ）	68
5.5	急ブレーキに影響する地点の要因の効果（全データ）	69
5.6	svm-score が上位の正例の語	69
5.7	svm-score が上位の負例の語	70
5.8	急ブレーキの識別性能（学習用データ・5 分割交差検証）	71
5.9	急ブレーキの識別性能（全データ・5 分割交差検証）	73

第1章

序論

1.1 研究背景

社会を形作る交通インフラのひとつに道路がある。道路は人や物の移動に広く用いられ、歩行者や自転車、バイク、さらに自家用車、バス、トラックといった自動車が通行している。さまざまな人や車両が同じ空間を共用しているため、多くの要因によって交通状況が左右される。要因は大きく2つに分類され、道路の幅員や車線数、交差点の信号制御パターンといった道路上に常に存在する要因のほか、交通事故や道路沿線施設でのイベントのような突発的かつ一時的な要因がある。これらの要因の組み合わせで発生する渋滞の予測や、目的地までの所要時間の予測は困難である。交通状況の把握や、渋滞発生の予測、目的地までの所要時間の予測を実現し、道路交通を改善するために情報技術を活用して多くのシステムが開発運用されている。それらのシステム群を総括し、人・道路・車両の間で情報をやり取りして道路交通の諸課題を解決するシステムを、高度交通システム (ITS: Intelligent Transport Systems) と呼ぶ [1]。

ITS に用いられる要素技術のひとつに、全地球測位システム (GPS: Global Positioning System) や準天頂衛星システム (QZSS: Quasi-Zenith Satellite System) に代表される衛星測位システムがある。このシステムを ITS に応用する際には、各車両に衛星からの信号を受信する車載機器を設置して定期的に位置を計測し、その情報を集約蓄積する。蓄積されたデータを解析することで、特定の時間帯の交通状況や、特定区間の所要時間を定量的に把握し、道路交通の改善に活用することができる。

本研究では、道路を利用する公共交通機関である路線バスに着目する。路線バスは多くの課題に直面している。代表的なものに「遅延の常態化」「利用客の減少」の2つがある。これらの課題は、様相や程度は違うものの、都市部と地方の両方において共通に見られるものであり、社会を維持する上で大変重要なものである。

第一の課題「遅延の常態化」であるが、路線バスは運行する道路の混雑や多客のために

恒常的に遅延が発生し、時間信頼性が低下している。路線バスは運行ダイヤが定まっているため、道路混雑等による遅延が時間的尺度によって「〇分遅れ」と明確に現れることもあり、利用客の間には「バスは遅れやすい」という印象が定着している現状がある。大規模な路線バス事業者ではバスロケーションシステムを導入して運行状況を利用客に示し、バス到着までの待ち時間や目的地までの所要時間の判断材料として活用できるようにしている。しかし、バス運行状況の情報はあくまでも過去から現在までの情報を示しているに過ぎない。乗車予定のバスがいつ自分の乗車バス停に到着するかは、バスの現在地からバス停までの道路状況に左右されるため、正確に知ることは難しい。さらに、利用客の関心事項である「乗車してから下車するまでの運行状況の見込みや目的地への到着時刻」も、乗車予定バスが乗車バス停に到来する時刻を知るのが難しいのと同様に、予測が難しい。

第二の課題「利用客の減少」は、特に地方において顕著である。人口減少やモータリゼーションにより路線バス利用客の減少が続いており、特に中小規模の路線バス事業者では路線の維持が困難となっている。不採算を理由として、運行本数の削減、路線の縮小、自治体が運営するコミュニティバスやデマンドタクシーへの転換が行われるか、あるいは路線廃止に至る例が多い。存続しているバス路線についても、地域内の交通手段を維持するために国や地方自治体により運行補助金が拠出されているが、サービスの長期的維持と利用客の利便性維持向上のためには運行効率化によるコスト削減が欠かせない。

これら2つの課題を解決し、利用客にとってより利用しやすい路線バスのサービスを実現するためには次の2つの要素が必要である。第一に、運行状況を定量的に把握することが必要である。従来は、運行状況の把握が部分的、あるいは定性的なものにとどまることが多かった。定量的かつ網羅的に運行状況を把握することで、サービス改善につながる要素を新たに発見できると考えられる。第二に、バスの遅延が利用客にもたらす不便やイメージ低下を払拭し、利便性を向上させることが必要である。遅延を少なくするか、遅延が発生しても先の見通しを立てやすくして、移動にあたっての時間信頼性を改善する必要がある。

また、路線バスほか自動車にとって重要な事項に安全運転がある。安全運転が損なわれる事態のひとつに交通事故がある。交通事故の当事者となった場合、利用客や運転士の生命にかかわる事態も発生しうる。安全運転の実現のためには、交通事故が起こりやすい条件や要因を事前に把握することが重要である。それらの条件や要因を念頭に置いて防衛運転を実施することにより、安定した運行と信頼性の向上につながる。

課題解決に利用可能なシステムに、バスロケーションシステムがある。先述の通り、バスロケーションシステムは大規模な路線バス事業者には広く導入されている。このシステムを利用すると、バスの位置情報履歴や運行状況履歴が収集でき、解析に活用できる。しかし、バスロケーションシステムは一般には高価なシステムであり、中小規模のバス事業者はコスト面から導入を断念したり、あるいは一度導入してもシステム更新に困難を抱え

たりする問題が生じている。岡村ら [2] は、とりわけ地方の路線バス事業者にとって、ランニングコストが大きくなるバスロケーションシステムの導入が困難であることを指摘している。そうした事業者向けには低いコストで導入・運用ができる位置情報収集システムや運行状況解析システムが存在することが望ましい。バスロケーションシステムのコストの点について、公共交通利用促進ネットワーク [3] によるバスロケーションシステムの状況調査 [4] では、複数のシステム事業者による中小規模路線バス事業者向けバスロケーションシステムの製品導入事例が示されている。これらのシステムの多くは、大規模な路線バス事業者が導入運用しているシステムと比べて、安価かつ手軽に導入できることを主張している。また、玉田ら [5] は、低コストなバスロケーションシステムを実現する手法として、スマートフォンにアプリケーションを搭載したものをバス車両に設置して、バスの位置情報と運行便に関する情報を毎分サーバに送信蓄積するシステムを提案し、実際にサービスを実施している。

また、課題解決にあたり、機械学習技術が大変有用であると考えられる。機械学習技術を適用すると、人間の手では時間がかかる大量のデータの自動的な識別分類や、様々な特徴量を利用した推論による将来予測が可能である。また、機械学習による予測や識別に寄与した特徴量を解析して、従来は発見できなかったような要因の組み合わせを得ることもできる。

これら数多くの研究や実用化されているシステムにおいて未だ達成されていないこと、および、未探求の事柄として、本研究では次のことに着目した。

まず、バス運行所要時間予測手法について、多くの先行研究はリアルタイムな情報を入力として要請しており [6, 7, 8, 9, 10]、リアルタイム性が下がるような情報を利用した場合であっても、リアルタイム情報がある場合と遜色のない予測精度を実現できる手法が未探求である。各システムの仕様として、バス車両に搭載した位置情報計測機器と、位置情報等の運行状況集積システムとの間で携帯電話回線を用いたリアルタイムな通信をする形で実装されている。この部分において、何らかの事情でリアルタイムな情報が得られない場合、あるいは、システムの仕様がリアルタイムに情報を送受信する仕組みでない場合に、リアルタイム性の低い過去の情報のみを用いた運行所要時間の予測をする手法を開発する余地がある。

次に、路線バス等自動車の安全運転に有益な情報を獲得するにあたり、先行研究では交通事故やヒヤリ・ハット事象が発生した要因について、発生地点に関連する要因の重要度を分析したものが存在した。しかし、地点と直接関連しない要因、たとえば時間的要因や天候のような外的要因の中から、有力なものを自動的に抽出する手法は、著者が調べ得た限りにおいては見つからなかった。例を示すと、バスの1年間の運行において急ブレーキによって危機を回避した事例が複数存在した時、急ブレーキ発生地点は解析によって明確化され、その付随要因もデータや現地の状況を観察することで明らかにできる。しかし、

地点の要因やそれ以外の付随要因の重要度を判定することは大変難しい。

1.2 研究目的

これらを踏まえ、本研究では次の 3 点を明らかにする。

まず、バスの運行所要時間予測に必要な特徴量を探求する。その発展形として、リアルタイムな情報が得られない状況下でも予測可能な手法を開発し、リアルタイムな情報を利用する手法と比較して、精度に遜色がないことを示す。最後に、安全運転につながる要素として「急ブレーキ」に着目し、位置情報履歴データの解析と機械学習の適用により、急ブレーキが発生する要因を発見する手法を構築する。また、位置情報履歴データを収集利用するにあたり、低コストでのデータ収集を実現するためのシステムを検証し、予測に必要な情報が取得できるようにするための構成手法を明らかにする。

これらの研究の重要性は、次のとおりである。第一に、低コストでの位置情報履歴データ収集手法を確立することにより、従来はコスト面でバスロケーションシステムが導入できなかった中小事業者においても、バスの運行状況を定量的に把握できるデータを収集できるようになる。また、そのデータを活用して、運行所要時間の予測という、利用客の利便性向上に有用なサービスを提供できるようになる。第三に、安全運転につながる未知の要素を発見することで事故を未然に回避でき、利用客や運転士の安全を守ること、サービスへの信頼性向上につなげることができる。

1.3 本研究の独自性と貢献

本研究の独自性および貢献を以下に示す。

- 長期間にわたる大規模な位置情報データセットを利用した。
 - 1 年間を通してのデータであるから、季節による変動を織り込んで解析や予測ができています。
 - 先行研究では、長期間の解析は数台だけ抽出して実験しているものが多く、大量の車で解析した研究は期間が短い。多くの場合は、少数の台数かつ短期間にとどまる。
- リアルタイム性に欠ける情報でも高い精度で運行所要時間予測を実現した。
 - 拠点収集型の位置情報収集システムでも、所要時間予測が十分に可能。
 - リアルタイムなデータ送受信に必要な部分のコストを抑えることにつながる。
- 機械学習を利用して、急ブレーキ発生時における地点以外の要因を発見した

- 地点に依存しない要因として，時間的要因が上位に来ていることを発見した．

1.4 本論文の構成

本論文の流れは次の通りである．第2章では，本研究の基礎となるデータセットとして位置情報データセットを2点紹介する．また，データセットに対して実施した基礎的な処理と，機械学習を実施した際の予測精度評価指標の定義を示す．第3章では，路線バスの運行所要時間予測に必要な情報を示し，その情報が取得できずリアルタイム性の低い情報を用いる場合でも，リアルタイムな情報を用いた場合と比べて遜色のない予測精度を実現する手法を示す．第4章では，本研究で利用している位置情報履歴収集システムを利用して，所要時間の予測精度を高めるために必要な設備の追加設置箇所を選定する手法を明らかにする．第5章では，安全運転につながる要素として急ブレーキを取り上げ，急ブレーキ事象を機械学習で識別抽出することで，その要因を明らかにする．最後に，第6章で本研究を総括し，今後の課題を示す．

第2章

導入

2.1 はじめに

位置情報履歴データは、道路を走行する各車両に車両の識別子と走行日時、その時点での緯度・経度を記録する車載機を搭載し、定期的にその記録データを収集蓄積したものである。列挙した4つの情報の他、車両の速度、加速度やその方向など、挙動に関する情報が付加されることもある。これらのデータは、一定の時間間隔や一定の距離間隔で記録収集され、日々蓄積される。長期間かつ多数の車両にわたって収集されるので、非常に大規模なデータとなる。これらの車両識別子・日時・緯度・経度の情報から、車両の位置情報特有の有用な情報を算出、抽出することで、車両の運転特性や、運行所要時間の予測が可能となる。

本研究においては、2つのデータセットを利用した。第3章および第4章においては、路線バス事業者の昭和自動車株式会社（本社・佐賀県唐津市。以下、事業者名を「昭和バス」と記す。）が、伊都営業所（福岡県福岡市西区）に配置して運行している一般路線バス車両のべ42台から収集した位置情報履歴データを利用した。詳細は2.5節で述べる。第5章における、機械学習を用いた自動車の急ブレーキの要因解析では、会津若松市が提供する「会津若松市オープンデータ（公用車・公共交通車両走行情報）」を利用した。詳細は2.4節で述べる。

2.2 位置情報履歴データ収集

位置情報履歴データの収集システムには、さまざまな実装手法がある。ここでは、車載機の観点と、車載機からデータを収集する手法の観点から、既存研究で利用されている手法と本研究で利用している手法を例示する。

2.2.1 車載機による分類

まず、位置情報を計測するための車載機の観点で見ると、バスの運行支援用に特化したシステムの位置情報取得機能を利用するもの [8, 11, 12]、スマートフォン端末など汎用の位置情報計測機能を持った機器を利用するもの [13, 5, 14]、ドライブレコーダーの位置情報記録機能を用いるもの、自動車向けの位置情報収集記録装置を用いるものがある。

2.2.2 収集方式による分類

計測された位置情報を収集する方法の点では、携帯電話回線を利用して、走行中リアルタイムに位置情報を送信するもの [8, 11, 12, 13, 5, 14]、特定の箇所に設置された受信機の近傍を通過した場合に位置情報を送信するもの、運行拠点で手動で位置情報の記録を回収するものが挙げられる。

2.2.3 各実装方式の長所と短所

これらの組み合わせにより構築されるシステムには、それぞれ長所と短所がある。一般に、車載機については、バス専用のシステムを利用する場合のコストが高くなると岡村ら [2] が指摘している。一方、スマートフォン端末などの汎用機器を利用した場合は、機器設置コストを抑えられるものの、バス車両特有の熱や衝撃へのハードウェア耐性や信頼性に懸念があることを伊藤ら [14] が示している。位置情報の収集方法の点では、携帯電話回線を利用する場合は、回線契約に関するコストと、データ送信に応じて発生するコストを考慮する必要がある。車両が複数台ある場合は、車両台数分の回線契約が必要となり、個々の端末のデータ送信が発生することを考慮する必要がある。特定の箇所に設置した受信機でデータを回収する場合は、車両と受信機間の通信のコストは0とすることができる。また、受信機からデータ蓄積処理システムへの送信コストも、受信機の数だけの回線契約となる。しかし、この方法ではリアルタイムな位置情報は得られず、受信機の傍を1回通過したあと、もう1回通過するまでの時間分、位置情報履歴の取得が遅れる。

本研究の方式: ETC2.0 システムの位置情報収集記録機能

本研究で用いる昭和バスの路線バス位置情報履歴データは、自動車向けの ETC2.0 システムに備わる位置情報収集記録機能を利用して収集している。本システムでは、車載機で計測した位置情報を車載機内に蓄積する段階で一定距離ごとになるように間引いて記録する。また、車載機に蓄積されたデータを、運行経路上に設置された路側機を通過する際に無線通信で転送する。この方式の長所のひとつは、自動車向けに開発・供給されている車

載機を利用するため，自動車搭載にあたっての耐久性の問題が解決済みであるという点にある。また，専用の無線通信を利用するため，車両と路側機との間の通信コストが0となり，通信回線の契約を路側機の台数分だけに抑えられる点も長所となる。一方で，車載機に蓄積された位置情報履歴データを収集する間隔が路側機の通過タイミングに依存する。また，データ取得間隔を短くするために路側機を数多く設置すると，コストが上昇する点が課題として挙げられる。

2.3 位置情報履歴データ変換，統計量および特徴量算出

位置情報履歴データを解析に利用するためには，元のデータを変換したり，他のデータと組み合わせたりして，有用な情報を算出することが重要である。位置情報をオンラインの地図情報サービスなど，他のデータやサービスと組み合わせる場合は，データ上の位置情報と，対象サービスが取り扱う位置情報の座標の基準である測地系を統一する必要がある。現在，日本で用いられている測地系は，世界測地系である ITRF（国際地球基準座標系）に基づく「日本測地系 2011」（JGD2011）である。しかし，昭和バスの路線バス位置情報データは，旧日本測地系（Tokyo Datum）により記録されているため，測地系変換の処理を行っている。

次に，道路の状況や路線バスの運行状況を算出したい場合は，交差点や道路，バス停などの位置と照合することで，位置情報の履歴が，どのような部分を走行しているデータであるかを明らかにする必要がある。つまり，車両の位置情報と，バス停などの特定の地点との間でデータの対応付けをしなければならない。また，対応付けられたデータをもとに，区間の所要時間や，バスの運行路線，遅延といった情報を求める必要がある。

これらのデータ変換を実施して，解析や機械学習に必要な特徴量を算出した。2.4 節と 2.5 節で，各データセットの概要，および実施したデータ変換処理の概要を述べる。

2.4 データセット 1: 会津若松市公用車・公共交通車両走行情報

本データセットは，会津若松市が保有する公用車，および市内循環バス『エコろん号』に搭載された車載機器により収集された位置情報履歴データである。本データセットは，会津若松市が運営するオープンデータカタログウェブサイト「会津若松市 DATA for CITIZEN」[15]，および，会津若松市と協力してデータセットを公開するバンプレコーダー株式会社が運営するウェブサイト「BumpRecorder Web」[16]に掲載され，クリエイティブ・コモンズ 表示 4.0 国際 [17] によりライセンスされ利用できる。データ詳細を表 2.1 に示す。また，データセットの書式を表 2.2 に示す。各車両の位置情報や加速度が

表 2.1 「会津若松市公用車・公共交通車両走行情報」データ諸元

項目	諸元
期間	2015.10.01 – 2018.05.15
対象車両	会津若松市公用車 5 台, 市内循環バス『エコろん号』1 台
データ管理者	会津若松市役所企画政策部情報統計課
データ数	263,035,146 件 (約 2 億 6,300 万件)
データサイズ	25.8 GiB

表 2.2 「会津若松市公用車・公共交通車両走行情報」データ書式

項目名	値の例
測定車両名	douroiijika
測定 ID	20151030_162045
測定日時	2015/10/30 16:20:45
測定日時 (UNIX 時間 + ミリ秒)	1446189648706
緯度 (JGD2011)	37.490131
経度 (JGD2011)	139.928786
測位誤差 [m]	14
左右加速度 [m/s ²]	0.846
前後加速度 [m/s ²]	-1.448
上下加速度 [m/s ²]	9.600

0.1 秒間隔で記録されているため、データ数が膨大である。また、上下加速度については重力加速度の影響は取り除かれていない。

2.5 データセット 2: 昭和バス伊都営業所管内路線バス位置情報履歴データ

昭和バス伊都営業所管内路線バス位置情報履歴データは、昭和バス伊都営業所に配置されている一般乗合路線バス車両 42 台に搭載した ETC2.0 車載機によって計測蓄積された位置情報履歴データである。データ詳細を表 2.3 に示す。ここで、元データからバス停の通過情報を算出し、バスの起点から終点までの 1 回の運行を 1 つのシーケンスと数えたときのシーケンス数も示す。また、データセットの書式を表 2.4 に示す。

本データセットには、伊都営業所が管轄する九州大学線、西の浦線、マリノア線の 3 路線、および各路線に係る回送便の位置情報が含まれている。本研究では九州大学線に

表 2.3 「昭和バス伊都営業所管内路線バス位置情報履歴データ」データ諸元

項目	諸元
期間	2019.04.01 – 2020.03.31
対象車両	伊都営業所配置の一般乗合路線バス車両 42 台
データ数	14,096,954 件 (約 1,410 万件)
データサイズ	992 MiB
シーケンス数	65,633 件

表 2.4 「昭和バス伊都営業所管内路線バス位置情報履歴データ」データ書式

項目名	値の例
車両 ID	1
走行時刻	2021-08-01 15:04:07
緯度 (Tokyo Datum) (値を 10 万倍して整数化)	3359997
経度 (Tokyo Datum) (値を 10 万倍して整数化)	13021360
速度 [km/h]	16
前後加速度 (挙動履歴データのみに表示) [m/s^2]	-0.42
方位 (挙動履歴データのみに表示)	南

表 2.5 九州大学線の運行本数の変化

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

かかわるデータを抽出して利用した。九州大学線の路線概略図を図 2.1 に示す。九州大学線は福岡市西区を運行する路線で、JR 九大学研都市駅から九州大学伊都キャンパスまでの約 4km を結んでいる。データセットの期間中における運行本数を表 2.5 に示す。

システムにより収集されたバス車両の位置情報は、データ処理サーバによって外れ値や異常値の除去が行われ、データベースに蓄積される。この位置情報履歴データを取得し、本研究で解決する課題への利用が可能となるよう、さらに情報の抽出・算出処理を行う。本データセットは、基本情報として車両 ID、時刻、緯度・経度、速度が記録される。これらに加えて、車両に急ブレーキや急ハンドルなど、一定以上の加速度変化が生じた場合に、加速度の方向と大きさが記録される。解析にあたり、車両の位置情報とバス停などの

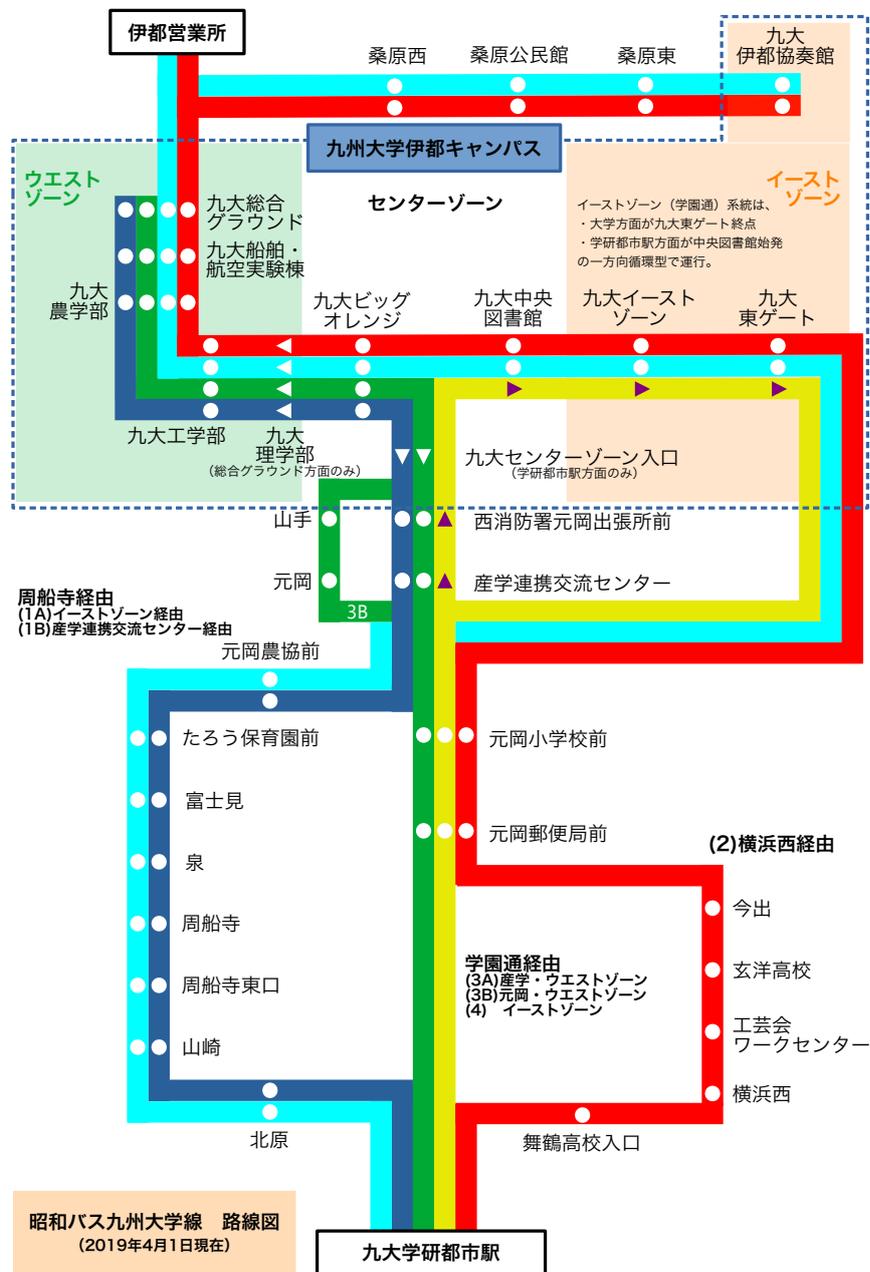


図 2.1 昭和バス九州大学線路線図 (2019 年 4 月 1 日現在)

特定の地点の情報の対応付けを行い、そのデータをもとに、バス停間の運行所要時間や、運行路線を求めた。

本データセットの位置情報は、緯度・経度の値が旧日本測地系 (Tokyo Datum) により記録されている。利用にあたり、地図やバス停の位置情報との照合を実施するために、世界測地系に基づく日本測地系 2011 (JGD2011) に従い、緯度・経度を変換した。本変

換には本来複雑な計算を必要とするが、国土地理院において、変換用プログラムのソースコード、および、変換パラメータファイルが公開配布されている [18]。このソースコードを国土地理院コンテンツ利用規約に従い、現在の手元環境で実行可能となるように変換したプログラムを作成した。プログラムを GitHub にて公開している [19]。

バス車両の位置情報履歴データに対して、その位置情報から一定の範囲内で最も近いバス停を求めて対応付けると、そのバス車両がいつバス停を通過したかがわかる。本データセットでは、対象とするバス停が 76 箇所、バス停位置を示す標柱が 151 本であったため、バスの位置情報の記録 1 件に対して、すべてのバス停標柱との距離を求めた。なお、計算を早期に打ち切って高速化を図るため、バスの位置情報レコードとバス停標柱間の距離が 50m 以下になった時に、当該バス停と対応付けて計算を終了することとした。距離計算にはヒュベニの公式の簡易式を用いた。本来の公式の理論は [20]、簡易式については [21, 22] に詳しい。距離 D を求める式は以下の通りである。

$$D = \sqrt{(Dy \cdot M)^2 + (Dx \cdot N \cdot \cos P)^2}$$

Dy : 2 点の緯度の差 (ラジアン単位)

Dx : 2 点の経度の差 (ラジアン単位)

P : 2 点の緯度の値の算術平均

$$M = \frac{Rx(1 - E^2)}{W^3}$$

$$N = \frac{Rx}{W}$$

$$W = \sqrt{1 - E^2 \cdot \sin^2 P}$$

$$E = \sqrt{\frac{Rx^2 - Ry^2}{Rx^2}}$$

M : 子午線曲率半径

N : 卯酉線曲率半径

E : 離心率

Rx : 長半径 (赤道半径)

Ry : 短半径 (極半径)

長半径と短半径の値は、JGD2011 において準拠する GRS80 楕円体により定義される [23]。値を表 2.6 に示す。

短半径の値は次のように求められる。

表 2.6 GRS80 楕円体の長半径と扁平率

長半径 R_x	6,378,137 m
扁平率 f	1/298.257222101

表 2.7 ルート別シーケンス数 (九州大学線)

経由	九州大学方面	九大学研都市駅方面	計
学園通経由	15,526	16,472	31,998
周船寺 (イーストゾーン) 経由	7,085	10,329	17,414
横浜西 (イーストゾーン) 経由	7,600	8,621	16,221
計	30,211	35,422	65,633

$$\begin{aligned}
 R_y &= R_x(1 - f) \\
 &= 6,356,752.314245 \text{ m}
 \end{aligned}$$

位置情報履歴とバス停位置との対応付けをした場合、1つのバス停に対して、複数の連続したバス位置情報の記録が対応付けられることがある。この場合は、そのデータ群の時刻情報のうち、もっとも早い時刻をバス停への到着時刻、もっとも遅い時刻をバス停からの発車時刻とみなした。また、到着時刻と発車時刻の間の経過時間を、当該バスの停車時間とした。このことにより、バスの所要時間を、各バス停間の所要時間と、各バス停での停車時間に分けることが可能となる。

バス停との対応付けが行われた位置情報履歴データは、バスの起点や終点となるバス停の部分でデータを区切ることができるようになる。起点から終点までの位置情報の点群データを1シーケンスと定義する。シーケンスへの分割にあたり、すべてのバスの起終点となるバス停と、一部のバスに対してのみ起終点となり、その他のバスがさらに先へと運行するバス停の両方を考慮して分割する必要がある。そこで、本研究では、位置情報履歴の記録日時と、それから算出したバス停着発時刻の情報をもとに、起終点の候補となるバス停で一定以上の時間が空いている場合は、シーケンスの分割点とみなした。各シーケンスに含まれるバス停情報のうち、各路線に特有のバス停を通過していることを検知して、そのバスの運行路線を決定する。本研究で対象とする路線は大きく3つのルートに分かれる。ルートにより分類したシーケンス数を表 2.7 に示す。

2.6 予測精度評価指標

本研究で実施した機械学習の分類性能や予測精度を測るために、次の指標を利用した。

表 2.8 混同行列

		推定クラス	
		正	負
真のクラス	正	TP (真陽性)	FN (偽陰性)
	負	FP (偽陽性)	TN (真陰性)

2.6.1 分類問題に対する評価指標

正解率 (Accuracy)

正解率 (Accuracy) は、分類問題に対する機械学習の評価指標のひとつで、次の式で表される。

$$Accuracy = \frac{True}{n} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.1)$$

ここで、 n はデータ数、 $True$ は正解数である。また、 TP, FP, TN, FN の意味は表 2.8 の通りに分類結果を集計したものである。表 2.8 は混同行列 (Confusion Matrix) と呼ばれる。

精度 (Precision)

精度は、機械学習の分類器がデータのクラスを正と推定したときに、真のクラスが正である割合のことである。式は次の通りである。

$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

この指標を利用すると、予測の確実性を判断することが可能であるが、単に推定時に正とする基準を厳しくするようなモデルでも精度が高くなる問題がある。

再現率 (Recall)

再現率は、真のクラスが正であるデータに対して、どの程度正であると判定できたかを示す割合である。式を次に示す。

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

この指標を利用すると、真のクラスが正であるようなデータを誤って負と判定してしまう割合を減らすことができる。しかし、単に Recall が高いだけでは、すべてのデータに

対して正と判定してしまう分類器を排除できない。

F 値

精度と再現率はトレードオフの関係にあるため、これらを共に高めることができる分類器を判定する指標として用いられるのが、精度と再現率の調和平均をとった F 値である。定義を次に示す。

$$F = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.4)$$

2.6.2 回帰問題に対する評価指標

回帰問題に対する機械学習器の予測精度を表す指標のうち、本研究では平均絶対誤差 (MAE) と平均絶対パーセント誤差 (MAPE)、および MAAPE を利用した。

平均絶対誤差 (MAE: Mean Absolute Error)

平均絶対誤差は、真の値と予測値の差の絶対値について、全データにわたる平均値を取ったものである。誤差の値を直接に知ることができるが、誤差の絶対量を表す量であるため、異なるデータ群に対する機械学習器の予測精度を比較する際、データ群のスケールが違う場合は、指標として用いない方が良い場合もある [24]。本研究では、スケールの異なるデータ群間の比較には MAE を用いず、2.6.2 節で述べる MAPE や、2.6.2 節で述べる MAAPE を利用した。

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

平均絶対パーセント誤差 (MAPE: Mean Absolute Percentage Error)

平均絶対パーセント誤差は、真の値と予測値の差の絶対値について、真の値の絶対値で割った値の全データにわたる平均値を取ったものであり、誤差の程度を示す相対的指標である。スケールの異なるデータ群間の比較に利用できるが、真の値に 0 が含まれる場合には指標が計算できなくなり、0 に近い値の場合は指標が高くなり過ぎてしまう問題がある。

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

MAAPE: Mean Arctangent Absolute Percentage Error

MAAPE は、真の値と予測値の差の絶対値について、その余接を計算してデータ数で割った、予測精度の相対的指標である [25]. 平均絶対パーセント誤差 (MAPE) に存在する、真の値に 0 が含まれる場合に MAPE が計算できない問題や、真の値が 0 に近い値の場合は MAPE の評価値が高くなり過ぎる問題を解決している.

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

第3章

バス運行所要時間の予測

3.1 はじめに

路線バスの課題のひとつ「遅延の常態化」の解決においては、バスの遅延を2種類に分類することが有効である。第一は時間的傾向によって起きているような、運行ダイヤの調整によって解決可能な遅延、第二は日時によって変わり、運行ダイヤに反映しきれないか反映すると不便を招く遅延である。本章では、後者の遅延に着目し、機械学習を利用して問題を解決する。

バスの運行所要時間を予測することは、乗客のバス停での待ち時間や不安を軽減することに有用である。また、目的地に到着してからの行動計画や、他の交通機関への乗り換えが計画通りにできるかどうかの判断にも役立つ。さらに、バスの所要時間や停車時間を正確に予測・推計することは、バス事業者がサービス向上を進める方策を考えるのに良い指標を得ることにつながる。例えば、イベント発生時、停留所での乗客の混雑を緩和するための臨時バスの運行や、ダイヤの調整などを判断することができる。バスの運行所要時間を予測する研究はさかんに行われている。

バス運行所要時間の正確な予測は、高度交通システム (ITS) の分野において、挑戦的な研究課題のひとつとなっている。バス運行データの取得 [26, 27], バス到着時間の予測 (研究 [8, 11] など多数), 予測結果の活用 [28, 29, 30] など、様々な観点から正確な所要時間予測を目指した研究が行われている。また、所要時間を予測するためのモデルも数多く提案されている。これらは、静的モデル、動的モデル、機械学習モデルの3種類に大別される。静的モデルの代表的なものとして、回帰分析やカーネル回帰がある。動的モデルの例は、カルマンフィルタリングが挙げられる。機械学習モデルとしては、Artificial Neural Network (ANN), k-Nearest Neighborhood (k-NN), Long-Short Term Memory (LSTM), Recurrent Neural Network (RNN), サポートベクターマシン (SVM), ランダムフォレスト (RF), 勾配ブースティング決定木 (GBDT) など、多くのアルゴリズムが採

用されてきた。また、2つ以上のモデルを組み合わせた予測モデルを提案する研究も多い。

これらの研究が大きな成果を上げているにもかかわらず、予測性能の向上にはまだ多くの課題が残されている。本章では課題を2つ取り上げ、その課題を解決する。

1つ目の課題は、まだ得られていない新しいデータを収集し、そのデータから予測性能を向上させる有効な特徴量を発見することである。このような特徴量の選択は、モデルの予測性能を向上させる上で最も重要な課題のひとつである。特徴量の選択はデータの特性に強く依存するため、データを収集する際の規模、ばらつき、経路の特徴などに注意を払う必要がある。

2つ目の課題は、リアルタイムな情報が得られない環境下において高い所要時間予測精度を維持する手法の開発である。運行所要時間の予測精度を高めるために、先行研究の大部分において要求されている入力情報として、リアルタイムな運行情報がある。本研究で用いるような特定の拠点通過時にデータを回収する方式では、情報をリアルタイムに収集できない可能性がある。こうしたシステムに有効な予測手法の開発や、予測精度の検討は先行研究において十分に行われていない。

これらを踏まえ、本章では、リアルタイムな情報が得られない環境下でも、高い精度でバスの運行所要時間を予測するための手法を示す。具体的には、機械学習アルゴリズムとして勾配ブースティング決定木 (GBDT) を利用し、位置情報履歴データから、指定のバス停間および路線の起終点間の所要時間を予測するモデルを構築し、また構築にあたり有用な特徴量を選定した。

本章の主な貢献は以下のようにまとめられる。

1. 40台以上の多数の車両に搭載された位置情報計測機器と路側機を用いて、最長3年間にわたり収集した路線バスの位置情報履歴データからの運行所要時間予測手法を検討した。
2. 位置情報履歴データから抽出した4つの特徴量を組み合わせる手法を提案し、GBDTを用いた予測モデルを構築した。
3. 特徴量の組み合わせの効果を評価した結果、予測対象区間までの区間の運行所要時間と、予測対象区間に到達するまでにバスが停止した回数の2つの特徴量を用いると、最も良い予測性能を示すことを明らかにした。
4. 一方で、実際に運用している位置情報履歴収集システムでは、最も良い予測性能を実現できる特徴量を取得利用できないため、その場合でも遜色のない運行所要時間の予測精度を達成する特徴量を選定した。

3.2 関連研究

本章に関連する研究を、機械学習アルゴリズム、および特徴量エンジニアリングに大別して議論する。

3.2.1 機械学習アルゴリズム

Bai ら [10] は、中国深センの市バスのための動的旅行時間予測モデルを提案し、複数のバス路線がある道路でのケースを扱っている。提案モデルは、SVR や ANN、カルマンフィルタリング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 や ANN モデルを用いてバスの基準走行時間を予測し、次に、最新のバス運行情報とカルマンフィルタリングを用いた動的アルゴリズムを用いて推定基準走行時間の両方を用いてバス走行時間を調整するものである。提案モデルの性能は、中国深センの道路から得た実データで検証された。データは、バス停の組の間の 3 つの共通区間を含む 5 つのバスルートから得たものであった。Bai らは、カルマンフィルタリング動的アルゴリズムを用いない、SVM と ANN モデルによる 3 つの区間の移動時間の予測結果と比較することで、提案モデルの検証を行った。Chen ら [31] は、第一に、ANN とフィードの位置と APC によって得られた旅行時間を用いて、与えられた時間帯、曜日、天候における旅行の時間点間のバス旅行時間を予測するモデル、第二に、カルマンフィルタによる動的アルゴリズムを用いて、バス位置情報に基づいて到着時間の予測値を分単位で調整するモデルを提案した。Sinn ら [32] は、リアルタイムの GPS 測定に基づくノンパラメトリックアルゴリズムを提案している。キーとなるアイデアは、バスの位置と、バス停への到着時間の間の依存関係を表現するために、カーネル回帰モデルを使用することである。提案アルゴリズムの性能は、アイルランドのダブリンの公共バス交通システムから得られた実データで評価されている。時間予測の範囲が 50 分の場合、アルゴリズムの予測誤差は平均で 10% 未満である。

山口ら [33, 34] 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。

本研究では、アルゴリズムとして GBDT [35, 36, 37] を使用する。アルゴリズムの選択根拠は、山口ら [33] の先行研究において、LR, ANN, SVR, RF, GBDT の 5 つの機械学習アルゴリズムを比較した中で、Friedman が提案した競争力があり、高いロバスト性

と解釈可能な手順で、回帰問題と分類問題の両方で高い性能を示す GBDT の優位性を見出していることに由来する。

3.2.2 特徴量エンジニアリング: 予測に有用な特徴量の調査

本研究の興味は、運行所要時間予測に有用な特徴量を探索することである。この点については、先に紹介した研究および本節で紹介する研究において、様々な候補が挙げられ、その有用性が主張されている。バス停間における運行所要時間情報は、ほぼ全ての研究で収集されており、予測モデルの構築に利用されている。また、バス停間の情報も同様にほぼ全ての研究で利用されている。また、1日のうち、主に朝・昼・夕・夜、あるいはピーク時・オフピーク時に分類される時間帯を利用する研究もある。また、区間内の移動時間を統計的に特定の形に変換し、それを特徴量として利用する研究もある。

Asら [38, 39, 40, 6] は、西日本鉄道株式会社（福岡県福岡市）から提供された、1か月間のバスプローブデータを用いて、隣接するバス停間のバス走行時間の予測方法を提案した。彼らは、主に所要時間が不安定な区間を対象とし、履歴平均移動時間（HATT）と動的平均移動時間（DAT）という2つの入力データを用いてバス移動時間の予測を行っている。HATT は、最近数日間の特定時間帯における、隣接するバス停のペア間の各区間の平均所要時間である。DAT は、同じ日の現在の時間帯の直前の時間帯の各区間の平均移動時間であり、ルート上の異なる日の間の変動に対処することが可能である。彼らは予測モデルの構築に ANN と SVR モデルを使用した。山口ら [33, 34] は、Asらと同じデータを用い、彼らより多い 80 以上の路線を扱い、隣接するバス停の組の各区間における時刻表上での所要時間と、実際の所要時間との差である、バス遅延時間の予測モデルを提案している。他の研究では、路線上の総所要時間 [41]、バス停の数と停車時間 [9]、バスの乗客数、バスを識別するためのバス ID や路線 ID を特徴量として追加している [42] などである。外的条件として最も利用されているのは、気象情報である [13]。道路混雑を特徴量として定量化した研究もある [43]。

田中ら [44] 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障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。

本章では、予測モデルの構築に用いる、予測対象区間の隣接するバス停の組と、予測対象区間の始点バス停の出発時刻という2つの基本的な特徴量に加え、以下の4つの特徴量を候補として考え、特徴量の組み合わせによって予測精度がどのように変わるかを検討する。

表 3.1 データの概要

項目	件数
バス車両台数	42
データ期間	2019.04.01 – 2019.12.31
位置情報履歴数	10,928,775
シーケンス数	191,785
(うち学習用データ)	138,182
(うちテストデータ)	53,603

- A) 予測対象区間の直前までの所要時間
- B) 予測対象区間に到達するまでのバスの停車回数
- C) 予測対象区間に到達するまでのバスの累積停車時間
- D) バスの運行曜日

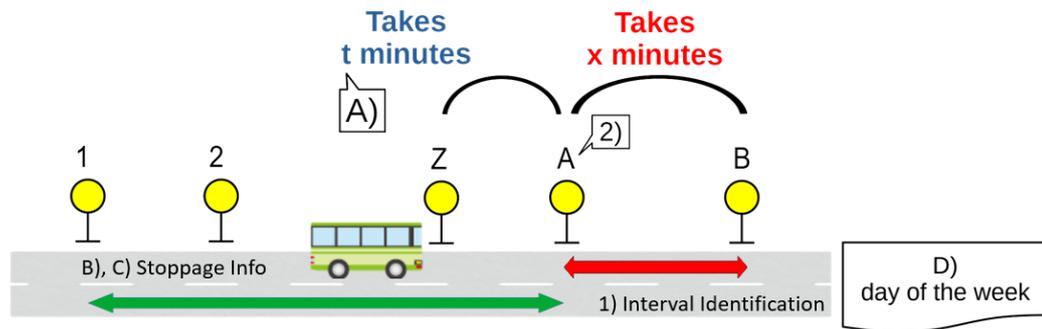
特徴量 A) と D) は, As ら [38, 39, 40, 6] が用いている. 本章において新たに特徴量 B) と C) を候補とした理由は, Huo ら [45] がバスの停車により遅延が増加し, 所要時間予測に影響を及ぼすと述べていることを考慮したためである. これら 4 つの特徴量の中から 1 つ以上の特徴量の組み合わせを作り, それを用いて予測モデルを構築する. 各区間の運行所要時間の予測性能に関する実験結果を評価する尺度として, 平均絶対誤差 (MAE) と平均絶対パーセント誤差 (MAPE), MAAPE[25] を採用する. その結果を, 特徴量の効果, すなわち, 特徴量が予測精度の向上にどの程度影響するかという観点で議論する.

3.3 データセット

本章のデータセットは, 昭和バス伊都営業所バス位置情報履歴データの一部である. 解析に利用した期間の詳細を表 3.1 に示す.

算出した特徴量は次の通りである.

- 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.



[Features to use]

- 1) The target interval between pairs of adjacent bus stops.
- 2) The departure time from the bus stop at the beginning of the target interval.

[Additional Candidate Features]

- A) Travel time over the interval just before the target one.
- B) The number of times a bus stopped before reaching the target interval from the starting bus stop on the route.
- C) The total stoppage time of a bus before reaching the target interval from the starting bus stop on the route.
- D) The day of the week on which the bus is running.

図 3.1 特徴量の概念図

3.4 実験手法

3.4.1 利用した特徴量

本研究で使用するデータは、以下の特徴量の組み合わせである。予測モデルに与える入力データとして、予測対象区間のバス停 ID の組（特徴量 1）と予測対象区間の始点バス停の出発時刻（特徴量 2）を常に使用している。この他に、特徴量 A)~D) の中から 1 つ以上の特徴量を選択して入力データとして使用する。予測モデルの出力は、入力で指定された区間における運行所要時間の予測値である。その概念図を図 3.1 に示す。

特徴量を以下に説明する。

1) 隣接するバス停の組の識別子

バス停 j とバス停 $j + 1$ の間を $I_{j,j+1}$ と表す。バス停数を m としたとき、 $j = 0, \dots, m - 1$ となる。各バス停の番号は、行きと帰りの重複がないように割り当てられている。

2) 予測対象区間の始点停留所の発車時刻

予測対象区間 $I_{j,j+1}$ の始点停留所 j における、UNIX Time で表す。

A) 予測対象区間の前の区間の所要時間

予測対象区間 $I_{j,j+1}$ の直前の区間 $I_{j-1,j}$ の所要時間と、予測対象区間 $I_{j,j+1}$ の所要時間には相関があると仮定する。区間 $I_{j-1,j}$ の所要時間を用いることで、予測対象区間 $I_{j,j+1}$ の所要時間の予測精度を向上させることができるはずである。

B) 予測対象区間に到達するまでの停車回数

バスの停車回数が増えると、バスの遅れがますます大きくなり、運転手が遅れを小さくするために急いで運転する可能性があると思定される。したがって、予測対象区間に到達するまでのバスの停車回数の情報は、予測対象区間の所要時間の予測精度に影響すると考えられる。

C) 予測対象区間までの総停車時間

B) と同様に、バスの総停車時間が長くなると、バスの遅れが大きくなり、運転手が遅延を小さくするために急いで運転する可能性がある。したがって、この情報も移動時間の予測精度に影響を与えると推測される。

D) 運行曜日

道路状況や混雑状況は、バス以外の自動車やバスの乗客の行動によって影響を受けると考えられる。そのため、曜日は、道路状況や混雑のばらつきを説明するのに有効な要素であり、それが所要時間に影響すると考えられる。ただし、祝祭日は日曜日と同じ時刻表が使われるため、祝祭日は日曜日と同等として扱う。

これらの特徴量を利用して、GBDT のアルゴリズムにより運行所要時間予測モデルを構築し、予測精度を検証した。予測精度の指標として、MAE と MAAPE を用いた。

3.5 結果

3.5.1 概要

実験結果を表 3.2 に示す。A~D 列は以下の特徴量を表し、使用する場合は 1 と記載する。

1. A: 予測対象区間の前の区間の所要時間
2. B: 目標区間に到達するまでのバスの停車回数
3. C: 目標区間までのバスの総停車時間
4. D: バスの運行曜日

1 行目 (Nothing) は、これら 4 つの特徴量を使用しない場合、つまり、学習とテストにおいて特徴量 A~D を使用しない場合の MAE の値である。この値を基準として、各特徴量を加えた場合の予測精度を評価した。表 3.2 に示すように、特徴量 A, B を追加することで最も良い MAE 値が得られた。すなわち、予測対象区間の前の区間の所要時間と、

表 3.2 予測精度評価

A	B	C	D	MAE[秒]	MAAPE			
					路線全区間	各区間平均	各区間最良値	各区間最悪値
(Nothing)				18.11	18.7%	19.0%	8.6%	41.9%
			1	18.38	19.1%	19.4%	8.7%	43.2%
		1		17.37	18.3%	18.2%	8.8%	39.2%
		1	1	17.71	18.6%	19.1%	8.5%	42.6%
	1			17.84	18.4%	18.2%	8.8%	37.7%
	1		1	17.74	18.6%	18.6%	8.6%	40.7%
	1	1		17.56	18.3%	18.4%	8.7%	38.4%
	1	1	1	17.58	18.1%	18.2%	8.8%	39.6%
1				13.42	14.3%	14.7%	3.9%	40.4%
1			1	13.14	14.1%	14.5%	3.4%	41.8%
1		1		12.39	13.0%	14.1%	4.0%	44.7%
1		1	1	13.81	14.9%	15.2%	3.7%	41.2%
1	1			11.66	12.3%	12.7%	2.9%	38.4%
1	1		1	11.88	12.4%	12.7%	3.2%	38.7%
1	1	1		12.02	12.8%	13.0%	3.1%	38.8%
1	1	1	1	13.35	14.6%	15.9%	3.8%	52.9%

予測対象区間に到達するまでのバスの停車回数が精度向上に影響を与えたといえる。特徴量 A, B を追加した場合の各区間の MAAPE の最悪値は、特徴量 B のみを追加した場合の値より良くはないが、基準より良い値となっている。表 3.2 では、特徴量 A を追加した場合、区間ごとの MAAPE の最悪値を除き、MAE および MAAPE の値が大きく改善されていることがわかる。また、特徴量 A と共に特徴量 B を追加すると、MAE, MAAPE ともにさらに改善された。一方、特徴量 C, D を追加した場合の改善効果は非常に小さく、MAE, MAAPE ともに値が悪化している場合もみられた。

3.5.2 各バス停間の予測精度の評価

表 3.2 に示した特徴量の追加による各バス停間での予測精度の変化を検証するため、区間ごとに MAAPE を評価した。図 3.2 は A, B, C, D の各特徴量を追加しない場合の MAAPE のヒストグラム、図 3.3 は特徴量 A, B を追加した場合の MAAPE のヒストグラムを示している。両者の間に有意差があるかどうかを統計的に検定した。MAAPE の

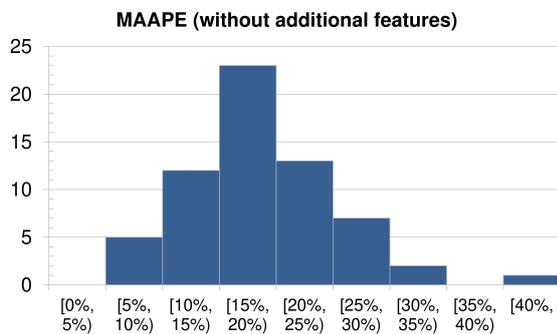


図 3.2 特徴量 A, B を追加しない場合の MAAPE

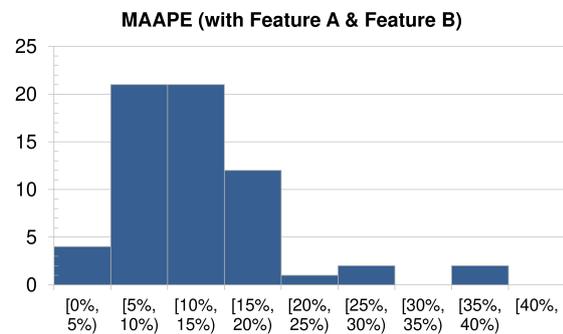


図 3.3 特徴量 A, B を追加した場合の MAAPE

表 3.3 特徴量と MAAPE との相関係数

特徴量 B	特徴量 N
0.20076	-0.05499

分布は正規性が仮定できないため、ノンパラメトリック検定手法の一つであるウィルコクソン符号順位和検定 [46] を適用した。ここで、検定の帰無仮説 H_0 は「両者の分布に有意差はない」である。その結果、 $p = 1.648 \times 10^{-11}$ となり、 H_0 は棄却された。つまり、特徴量 A, B を追加することで MAAPE が有意に改善されたと言える。

3.5.3 予測精度に影響を与えた特徴量の調査

特徴量 A と B を追加した場合に、最も所要時間の予測誤差が小さくなることがわかった。一方、特徴量 B のみでは、予測誤差を小さくする効果は弱い。したがって、特徴量 B の特性、特にデータ数との関係を詳しく調べる必要がある。各区間の運行所要時間を算出するためのデータ数を特徴量 N と定義する。ここで、2 つの特徴量 B と N が MAAPE に影響を与えるかどうかを確認するために、MAAPE とそれぞれの特徴量 B と N の相関係数を計算した。表 3.3 に示すように、MAAPE と特徴量 B, N との相関は、それぞれ非常に弱いとわかった。

特徴量 B と MAAPE との相関は、特徴量 B による運行所要時間予測誤差の改善効果が弱いことから、相関係数と予測誤差の改善効果の度合いがよく一致していると考えられる。しかし、特徴量 N は MAAPE と強い相関があると予想されたため、特徴量 N と MAAPE との相関が非常に弱いのは意外であった。

3.6 理想的な特徴量が得られない場合

前節で、運行所要時間の予測に最も有効な特徴量を明らかにすることができた。ここで、この結果は、すでに取得変換済の9か月間にわたる位置情報履歴データを一括してデータセットとし、解析や機械学習を実施したものである。解析結果は、運行所要時間の予測にはリアルタイムな情報が必要であることを示唆している。しかし、本研究で利用している位置情報履歴収集システムでは、位置情報履歴の収集は九州大学伊都キャンパス構内入口1箇所のみ設置された路側機を通過したタイミングに限られるため、リアルタイムな情報を得ることができない。また、運行所要時間の予測のためにはデータの変換が必要となり、処理に時間を要することから、実質的にはリアルタイムな情報の活用が困難である。さらに、前節では運行所要時間の予測対象区間をバス停間1区間に限っていた。実際の運行所要時間予測では、起点から終点までの間に複数のバス停が存在することが多い。この場合にも、予測に有効とされた特徴量を取得できない場面が生じる。たとえば、予測対象区間の始点が路線自体の起点である場合、「予測対象区間の1区間前」は存在しない。

このため、本節では、理想的な特徴量が得られない場合に代わりに取得・算出可能な特徴量を議論し、その特徴量を利用した場合の予測精度の変化を解析する。本節での運行所要時間の予測対象区間は、起点が九州大学（九大総合グラウンドバス停）、終点が九大学研都市駅である。

3.6.1 予測に利用可能な特徴量の条件

本研究の位置情報履歴データ収集システムの場合はリアルタイムな情報の取得は困難であるため、利用可能な特徴量は、予測対象のバス便が予測対象区間の始点バス停を出発する時点ですでに取得できているものに限られる。第4章で詳細に議論するが、リアルタイムではない過去の情報であれば取得できていると考えられるため、その情報を利用する。本節では、過去の情報として「予測対象区間における前の時間帯の運行所要時間の平均値」を利用した。この情報はAsら[47]が利用している。前の時間帯の定義として、本節では1時間前を例にとり、解析を実施した。この特徴量を選定採用した根拠は、運行所要時間を直接表している特徴量であることと、本研究の対象とする昭和バス九州大学線において、バスの時刻表上の運行所要時間が約30分とされていることから、収集および処理に要する時間等も加味しつつ余裕を取った時間設定としたことに由来する。

表 3.4 予測精度比較 (MAE・単位は秒)

パターン	経由	区間所要時間を加えない	大学出口まで加える	終点まで
A	学園通	94.03	91.68	61.81
	周船寺	108.85	108.28	77.51
	横浜西	93.65	80.26	57.24
B	学園通	69.13	60.17	47.22
	周船寺	52.55	49.85	44.57
	横浜西	39.69	33.73	31.85

表 3.5 予測精度比較 (MAAPE)

パターン	経由	区間所要時間を加えない	大学出口まで加える	終点まで
A	学園通	8.99%	8.43%	5.74%
	周船寺	6.48%	6.43%	4.65%
	横浜西	6.48%	5.60%	4.02%
B	学園通	6.57%	5.75%	4.51%
	周船寺	3.25%	3.06%	2.74%
	横浜西	2.77%	2.39%	2.26%

3.6.2 解析結果

2019年度の位置情報履歴データをもとに、各ルートにおいて予測対象区間の1時間前の運行所要時間平均値を利用して予測した場合の精度を解析した。結果を表3.4、および表3.5に示す。ここで、大学出口のバス停は、学園通経由の場合は九大ビッグオレンジバス停、周船寺経由および横浜西経由については九大東ゲートバス停を指す。また、パターンAは1時間前の運行所要時間の特徴量として利用しない場合、パターンBは利用する場合を意味する。表3.4、および表3.5の最終列には、予測対象便の予測対象区間終点までの各バス停間の所要時間をすべて特徴量として加えた場合の数値を参考として掲載している。これは実際の予測上は無意味な値である。その理由は、予測対象区間終点までのバス停間運行所要時間が得られた時点で、そのバスはすでに終点に到着しているためである。

解析の結果、1時間前の運行所要時間の特徴量として利用すると、すべての経由において予測精度が改善することがわかった。また、その改善度は、予測対象便自体のバス停間運行所要時間を大学の出口など途中の区間まで特徴量として加えて、予測対象区間の着点までの所要時間を予測する場合よりも大きかった。

表 3.6 特徴量重要度（パターン A・学園通経由・区間所要時間を加えない）

順位	特徴量名	重要度	説明
1	Hour	41.5	起点出発日時（時）
2	Month	29.5	起点出発日時（月）
3	Day	29.0	起点出発日時（日）

表 3.7 特徴量重要度（パターン A・学園通経由・大学出口まで加える）

順位	特徴量名	重要度	説明
1	Hour	22.6	起点出発日時（時）
2	Month	14.0	起点出発日時（月）
3	Day	13.3	起点出発日時（日）

表 3.8 特徴量重要度（パターン B・学園通経由・区間所要時間を加えない）

順位	特徴量名	重要度	説明
1	d_1hour_before	57.2	1時間前の運行所要時間の平均値
2	Hour	15.1	起点出発日時（時）
3	Day	14.9	起点出発日時（日）

表 3.9 特徴量重要度（パターン A・周船寺経由・区間所要時間を加えない）

順位	特徴量名	重要度	説明
1	Hour	36.1	起点出発日時（時）
2	Month	31.9	起点出発日時（月）
3	Day	31.7	起点出発日時（日）

3.6.3 1時間前の運行所要時間の特徴量としての重要性

GBDT を利用した機械学習の際には、どの特徴量が予測に寄与したかを示すスコアとして特徴量重要度（Feature Importance）が得られる。この値をもとに、有用であった特徴量を調査した。各パターン・経由による特徴量重要度の上位 3 件を抽出して示す。

これらの特徴量重要度の結果から、1時間前の運行所要時間の平均値は極めて有用であると言える。また、この特徴量が無い場合は、バスの起点出発日時のうち時の部分が特徴量として常に有効である。

表 3.10 特徴量重要度（パターン A・周船寺経由・大学出口まで加える）

順位	特徴量名	重要度	説明
1	Hour	14.9	起点出発日時（時）
2	Month	12.6	起点出発日時（月）
3	Day	10.6	起点出発日時（日）

表 3.11 特徴量重要度（パターン B・周船寺経由・区間所要時間を加えない）

順位	特徴量名	重要度	説明
1	d_1hour_before	55.5	1 時間前の運行所要時間の平均値
2	Day	15.4	起点出発日時（日）
3	Month	14.9	起点出発日時（月）

表 3.12 特徴量重要度（パターン A・横浜西経由・区間所要時間を加えない）

順位	特徴量名	重要度	説明
1	Hour	38.6	起点出発日時（時）
2	Day	33.2	起点出発日時（日）
3	Month	28.2	起点出発日時（月）

表 3.13 特徴量重要度（パターン A・横浜西経由・大学出口まで加える）

順位	特徴量名	重要度	説明
1	Hour	19.0	起点出発日時（時）
2	Month	11.0	起点出発日時（月）
3	d_47to45	13.3	所要時間（九大イーストゾーン→九大東ゲート）

3.7 先行研究との予測精度の比較

3.6 節で得た運行所要時間の予測精度が、先行研究と比較してどの程度良いかを調査した。各研究および本研究は対象とする地域や路線数、路線の長さ、データの期間、アルゴリズムなど様々な条件が異なっている。そのため、単純な比較が困難であるものの、異なる条件の研究成果を比較する際に有効な相対的指標を用いることとする。先行研究では予測精度の相対的指標として平均絶対パーセント誤差（MAPE）を採用しているため、本研究の成果についても、MAPE を算出して比較する。予測精度の概要を表 3.15 に示す。

表 3.14 特徴量重要度 (パターン B・横浜西経由・区間所要時間を加えない)

順位	特徴量名	重要度	説明
1	d_1hour_before	61.9	1時間前の運行所要時間の平均値
2	Hour	13.7	起点出発日時 (時)
3	Day	13.4	起点出発日時 (日)

表 3.15 先行研究の予測精度との比較 (MAPE)

研究	MAPE	算出条件
本研究	6.6%	1時間前の所要時間利用の場合の最悪値
Gurmu ら [48]	18.3%	平均値
Yu ら [30]	6.7%	平均値の最良値
Liu ら [49]	4.0%	長距離対象の場合の最良値
Bai ら [10]	4.3%	短距離の最良値
Pang ら [42]	14.4%	最良値
He ら [50]	4.1%	最良値
内村ら [8]	14.3%	最良値

先行研究はいずれもリアルタイムな情報、リアルタイムに近い情報が得られている状況において成立するものである。一方、本研究はリアルタイム性が低い情報を利用しつつも、全てのパターンにおける最悪値をもって比較して、先行研究と遜色のない予測精度を達成していると言える。

3.8 考察

3.8.1 特徴量の効果

運行所要時間の予測にあたって取得可能な特徴量は、予測対象区間の前にバスの運行区間があるかどうかで変わってくる。予測対象区間の前にバスの運行区間が存在する場合は、次の特徴量が有効であった。

- バスがどのバス停間を運行しているか。
- 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

予測対象区間がバス停1区間である場合は、予測対象区間の前の区間の所要時間（特徴

量 A) を特徴量として追加すると MAAPE が改善された。特徴量 A が予測精度の向上に寄与した理由としては、同じバス運行時間帯において、予測対象区間の前の区間を含む周辺の道路状況が類似しているためと推測される。一方で、特徴量を追加することにより予測性能が低下する場合が存在することが分かった。そのため、精度の高い予測モデルを構築するためには、特徴量を慎重に選択する必要がある。

ここで、予測対象区間に到達するまでの特徴量は、予測対象区間の始点がバス路線の起点であったような場合には取得できないため、「運行所要時間の予測対象区間における、1時間前の所要時間実績の平均値」を新たな特徴量として算出し、機械学習による予測を実施した。

3.8.2 予測誤差を生じる要因の調査

そこで、予測誤差に影響を与える3つの要因について検討した。1つ目の要因は、隣接するバス停間の距離であり、距離が短いほど、すなわち所要時間が短いほど、各バス停での停車時間や滞留時間の影響を強く受ける。そこで、本章では、バス停での停車時間を、そのバス停を起点とする区間の運行所要時間を含めることとした。これは、運行所要時間が短いほど、その区間の出発バス停での停車時間の分散に応じた運行所要時間の分散が大きくなることを想定したものである。しかし、残念ながら、この特徴量は運行所要時間の予測性能を向上させることに寄与せず、表 3.2 に示すように、最良の予測性能を達成するのにも効果がなかった。

もう一つは、バス停以外でのバスの停車時間、例えば信号で停車する回数が、区間での運行所要時間に影響し、運行所要時間に対する総停車時間の比率の分散が大きくなると予測誤差が大きくなることである。ここで、バスの停車時間は、途中のバスの停車回数にも影響を受けると考えることができる。この要因は、3.5 節の特徴量 A (対象区間の直前の区間の運行所要時間) と組み合わせることで、所要時間の予測性能の向上を図る効果がある。

3つ目は、各区間の旅行時間データの数を示す特徴量 N の効果を検証したところ、特徴量 N と所要時間予測精度の MAAPE との相関はほとんどないことが確認された。

3.8.3 1時間前の運行所要時間の効果

表 3.4 と表 3.5 で示したパターン A とパターン B の結果を比較すると、特徴量「1時間前の予測対象区間の運行所要時間の平均値」を利用した場合に、予測精度が大きく改善された。また、その改善度は、予測対象便そのものの途中までの運行実績所要時間を特徴量として加えて、予測対象区間の着点までの所要時間を予測する場合よりも大きかった。こ

のことは、リアルタイム性が低い過去の情報を利用して予測する場合でも、特徴量を適切に選択することにより、リアルタイムな情報を利用するのと遜色のない精度で所要時間が予測できることを意味していると考えられる。

リアルタイム性が低い情報ながら予測精度を高くできたのは、特徴量が反映している情報の差のためと考えられる。予測対象便そのものの途中までの運行実績所要時間の情報は、リアルタイム性が高い一方で、特徴量として入力した区間より先の情報を含んでいない。一方、1時間前の予測対象区間の運行所要時間の平均値は、過去の情報ではあるが予測対象区間の全区間の情報を含んでいる。また、当路線において時間帯による周期的な運行所要時間の変動があり、その周期性のため過去の情報から高い精度で運行所要時間を予測できていると考えられる。

3.9 まとめ

本章では、バス停間の運行所要時間を予測する方法を明らかにした。まず、9か月にわたる位置情報履歴データを一括してデータセットとして取り扱い、運行所要時間予測に有効な特徴量を解明した。予測モデルは、データセットから得られる特徴量 A：予測対象区間の前の区間の所要時間、B：予測対象区間に達するまでのバスの総停車回数、C：予測対象区間に達するまでのバスの総停車時間、D：バスが走行する曜日と、予測対象区間の始点バス停での出発時間から成る特徴量の組み合わせを利用して構築した。

解析の結果、運行所要時間の予測精度の向上には、特徴量 A および B が効果的であることがわかった。その他の特徴量を追加した場合の予測精度の改善効果は低く、逆に精度が低下する場合も見られた。各区間の運行所要時間予測値の MAAPE を比較すると、MAAPE の値には大きな差があり、最も悪い MAAPE は約 40% であった。また、各区間の運行所要時間のデータ数（特徴量 N）と MAAPE の相関係数を求めたが、ほとんど相関がなかった。

これらの結果からは、運行所要時間の予測にはリアルタイムな情報が必要であることが示唆されるが、本研究で利用している位置情報履歴データ収集システムでは、実際にはそれらの情報を取得できない。また、予測対象区間を隣接バス停 1 区間から複数のバス停間にまたがるような場合に拡張したとき、予測対象区間に到達するまでの特徴量という、予測対象区間の始点がバス路線の起点であったような場合には取得できないものが含まれる。そのため、「運行所要時間の予測対象区間における、1時間前の所要時間実績の平均値」を新たな特徴量として算出し、機械学習による予測を実施した。この特徴量はリアルタイム性の低い情報であるが、他にリアルタイム性の高い情報を加えずとも、先行研究と遜色のない運行所要時間予測精度を達成した。

次章では、複数のバス停区間から成る区間の運行所要時間を予測するにあたり、特徴量

として見出した「予測対象区間に到達するまでの所要時間」と、本章で新たに見出した「運行所要時間の予測対象区間における、1時間前の所要時間実績の平均値」が実際に取得可能であるかを解析する。取得可能であるならば、位置情報履歴データ収集用の路側機をどのように配置すればよいかを解析する。

第4章

低コストと高予測精度を実現する データ追加収集箇所選定手法

4.1 はじめに

本章では、路線バスの運行所要時間を予測するための位置情報履歴データを収集する設備について、低コストかつ高い予測精度を実現するためのシステム構築方式および収集用路側機設置箇所の選定手法を示す。具体的には、高い精度で運行所要時間を予測するのに有効とされた特徴量「予測対象区間に到達するまでの所要時間」と、本章で新たに見出した「運行所要時間の予測対象区間における、1時間前の所要時間実績の平均値」が実際に取得可能であるかを解析する。取得可能であるならば、位置情報履歴データ収集用の路側機をどのように配置すればよいか、また、その情報を収集するシステムの構成がどうあるべきかを示す。

路線バスの運行状況の把握や、運行所要時間の予測に利用するデータは、状況を即時把握できるようにリアルタイムで収集できたほうが望ましい。バスの運行所要時間や遅延を予測する多くの先行研究においても、予測精度を向上させるためには、リアルタイムな情報が最も効果的であるとの結果が得られたとの報告が多い。しかし、そうしたシステムは導入コストや運用コストが高くなるという課題がある。本研究で運用しているバス位置情報履歴データ収集システムは、車両に蓄積された位置情報データの履歴を、特定の拠点に設置した路側機を通過した時に回収する仕組みとしている。この場合、リアルタイムにデータを得るには、路側機の設置箇所を多くしなければならないため、路側機の数だけコストが上昇してしまう。そのため、リアルタイム性の低い情報でも高い予測精度を実現するための特徴量2点を実際に取得利用するために、路側機を設置する箇所として最適な場所を、機械学習を利用して調査した。

ここで、データ収集コストを小さくするためには、路側機の設置数を最小限度にすべき

である。まず、すでに1台の路側機が設置されていることを前提とし、その路側機1台で必要な特徴量を算出するための位置情報履歴データを得られるかを調べる。もし路側機を追加設置する必要があるならば、その台数をまず1台とし、運行所要時間の予測精度向上効果が最大となる地点に設置したほうが良い。加えて、バスの運行は往路と復路の両方向があるため、往路と復路の予測精度をともに向上させられる地点であることが望ましい。そこで、位置情報履歴データの解析により、路側機を追加設置地点を求めた。解析にあたっては、次のような場所を候補として検討した。

路側機増設地点の候補の1つ目として、バスの運行所要時間変動に大きな影響を与える区間の所要時間をいち早く得られるような場所が考えられる。そのような区間は、バス便により所要時間に大きな差が出ると考えられ、バスの運行所要時間の分散の値が大きくなると予想される。また、候補の2つ目は、運行所要時間予測対象区間における1時間前の所要時間実績値をを確実に得られる場所が考えられる。そこで、実際にある場所に路側機を追加設置したと仮定した場合に得られる位置情報履歴をもとに機械学習を行い、運行所要時間の予測精度を求め、その位置が当該ルートにおけるバス停間運行所要時間の分散の値が最大となっている区間と一致するかどうかを調べた。

4.2 関連研究

4.2.1 先行研究で必要とされてきたデータのリアルタイム性

従来より、路線バスの運行履歴データを用いたバス運行所要時間の予測は広く行われてきた。その際のデータ収集手法は、GPS等の衛星測位システムによって計測した位置情報やバス停の通過情報をリアルタイムに送信するもの、またはそれらの情報を路線中の拠点において収集するものの2通りに大別される。前者の例は内村ら[8]と今井ら[9]の研究など、多くの研究で見られる。両研究では高精度でバス運行状況を予測する手法を提案実証しているが、ともに予測に必須な入力情報として、リアルタイムな運行状況が欠かせないことを示唆している。Baiら[10]も、過去の運行履歴をもとに機械学習モデルを作成して、基礎となる運行所要時間予測値を算出する一方、最終的な予測結果の精度を高くするためには、リアルタイムな情報により予測値を動的に補正することが必要であることを示した。山口ら[33]、および川谷ら[7, 34]も、予測精度を高くするために必要な情報として、予測したいバス停間の1つ前の区間の運行所要時間が、機械学習による予測精度の向上に有効であることを見出し、リアルタイムな、あるいはできる限りリアルタイムに近いバスの運行所要時間情報が重要であるとの結果を得ている。

しかし、リアルタイムな情報を取得するためには、設備に対するコストや通信にかかるコストが大きくなる。岡村ら[2]は、とりわけ地方の路線バス事業者にとって、ランニン

コストが大きくなるバスロケーションシステムの導入が困難であることを指摘している。低コストなバス運行履歴データの収集手法として、玉田ら [5] は、スマートフォンにアプリケーションを搭載したものをバス車両に設置して、バスの位置情報と運行便に関する情報を毎分サーバに送信蓄積するシステムを提案し、実際にサービスを実施している。玉田らのシステムをはじめとする、リアルタイムに情報を送信する手法は、データ通信を行うにあたり、定期的に一定のコストが発生することは避けられない。

4.2.2 利用可能なデータの種類の制限がある場合

Maiti ら [51] は、シンプルで軽量な履歴データベースのモデルを提案し、非常に限られた情報（バスの位置と記録日時）しか入力特徴量として利用できない状況で、予測精度と予測時間のバランスをとる提案手法の可能性を示した。バスデータは4か月に渡って収集されている。

これらを踏まえて、本章では、バスの位置情報を得るために必要なコストを抑えたシステムを構築し、リアルタイムな情報がない状況においても、過去の情報と機械学習手法を用いて、高精度なバス停間運行所要時間を予測するために必要な、運行情報の取得箇所選定手法を提示する。

4.3 問題の定義

本章で解決する問題を2つ定義する。第一に、バスの運行所要時間変動に大きな影響を与える区間の所要時間をいち早く得られるような場所を発見する問題を定式化する。路線の起点 O から終点 D までの区間 $O-D$ 間に、バス停が $B_1, B_2, \dots, B_{N-1}, B_N$ と設置されているとする。この区間中に走行記録データを収集する路側機を設置する地点を1箇所定める。運行所要時間の予測精度を高くし、かつデータ収集コストを低くすることができる路側機の設置箇所を検討する。ここで、バス停 B_t と B_{t+1} 間に路側機 A を設置すると仮定する。この路側機 A においては、区間 $O - B_1 - B_2 - \dots - B_{t-1} - B_t$ の運行所要時間、および各バス停での停車時間が得られる。

第二に、運行所要時間予測対象区間における1時間前の所要時間実績値をを確実に得られる場所を発見する。まず、路側機が1台だけ設置されている現在の状況において、1時間前の所要時間実績値を得られるかどうか、バスの運行実績を統計的に解析して求める。もし、追加設置が必要であるならば、路線バス車両が既設の路側機を通過した後、1時間以内に通過する場所に設置すべきである。このような場所をデータから見出す。本章では、第一の問題に対する解析を「解析 I」、第二の問題に対する解析を「解析 II」として説明する。

表 4.1 データセット概要

項目	諸元
バス車両台数	42
データ期間	2019.04.01 – 2020.03.31
位置情報履歴数	14,293,440
シーケンス数	170,841

4.4 データセット

本章のデータセットは、昭和バス伊都営業所バスプローブデータの一部である。データセットとして採用した特徴量は解析パターンにより異なる。採用した部分のデータセットの詳細を表 4.1、また、利用した特徴量を表 4.2 に示す。

4.5 解析手法

4.5.1 データからのバス停間運行所要時間算出手法

データセットから、各バス停に停車、もしくは通過した時刻を求めた。簡略化のため、1 シーケンス中に、各バス停について、通過情報が 1 件現れた場合はその時刻を通過時刻とし、複数件現れた場合は、その中で最も早い時刻を到着時刻、最も遅い時刻を出発時刻とみなした。その上で、あるバス停 A の出発時刻 t_A とあるバス停 B への到着時刻 t_B の差、 $t_B - t_A$ をバス停 A-B 間の運行所要時間とした。

4.5.2 統計的解析

次に、各バス停間の運行所要時間について、基本統計量を求めた。統計量算出にあたり、次の基準から逸脱した値を外れ値として除去した。まず、データの不整合により所要時間が負の値となっているもの、所要時間とバス停間の距離から算出した移動速度が 60km/h を超えているとされたデータを除去した。次に、各バス停間の運行所要時間が極めて長くなっているデータを選別する基準として、基本統計量から求めた標準偏差 σ を用いて、各バス停間ごとに、所要時間平均値 $\mu \pm 3\sigma$ の範囲外の値を除去した。

表 4.2 解析に利用した特徴量

特徴量	解析 I	解析 II
運行日時（年）	○	○
運行日時（月）	○	○
運行日時（日）	○	○
運行日時（時）	○	○
天候	○	○
所要時間実績値（あるバス停間 1 区間のみ）	○	
所要時間実績値（起点からあるバス停まで）	○	
所要時間実績値（1 時間前の起点から終点まで）		○

4.5.3 機械学習

運行所要時間予測には、勾配ブースティング決定木（GBDT）を用いた。また、予測精度の評価指標は、MAE（平均絶対誤差）と MAAPE を利用した。機械学習器に対してデータから算出された各種の特徴量や、時刻情報との照合で得られる天候等の外部条件を加えて、バス路線の起点から終点までの運行所要時間を予測した。解析 I と解析 II で利用した特徴量を表 4.2 に示す。

ここで、運行日時は、起点バス停を出発した日付時刻を示す。また、天候は、福岡管区気象台において 1 時間ごとに観測公表されている天候をカテゴリ変数として扱ったものを加えた。あるバス停間 1 区間の運行所要時間は、1 つのシーケンスについて各バス停間の移動に要した時間、すなわち 4.5.1 節に示した $t_B - t_A$ である。これを 1 区間ずつ入れ替える形と、起点側から順次追加する形の両方で特徴量として用いた。1 時間前の運行における起終点間の運行所要時間は、運行所要時間の予測精度を改善するために用いる特徴量である。

また、GBDT では、各特徴量が予測にどのような影響を与えたかについて、その寄与度を示す特徴量重要度（Feature Importance）という指標が得られる。この指標を算出して、予測精度との関係を調査した。今回用いた GBDT ライブラリ CatBoost では、Feature Importance として PredictionValuesChange という値を算出する [52]。この値は、ある特徴量の変化が出力結果に与えた影響度をスコア化し、そのスコアの合計が 100 となるように正規化したものである。

表 4.3 運行所要時間の分散が大きい区間（九州大学方面）

路線	区間	分散 σ^2
学園通	九大学研都市駅→元岡郵便局前	3,553.16
〃	西消防署元岡出張所前→九大ビッグオレンジ	946.26
周船寺	九大学研都市駅→北原	1,492.57
〃	元岡農協前→九大東ゲート	1,173.73
横浜西	九大学研都市駅→舞鶴高校入口	2,184.71
〃	元岡小学校前→九大東ゲート	2,022.06

表 4.4 運行所要時間の分散が大きい区間（九大学研都市駅方面）

路線	区間	分散 σ^2
学園通	元岡郵便局前→九大学研都市駅	3,394.38
〃	産学連携交流センター→元岡小学校前	808.07
周船寺	九大東ゲート→元岡農協前	1,432.92
〃	泉→周船寺	1167.40
横浜西	舞鶴高校入口→九大学研都市駅	2,127.23
〃	九大東ゲート→元岡小学校前	1,103.76

4.6 解析 I: 所要時間予測精度を向上させる区間と当該区間の所要時間の分散の関係

4.6.1 あるバス停間 1 区間の所要時間のみを特徴量として加えた場合

既設の路側機の設置位置を考慮し、九州大学伊都キャンパス構内のバス停間については路側機追加の候補から除外した上で、運行所要時間の分散が大きい区間として、表 4.3 および表 4.4 の区間を見出した。

表 4.3 および表 4.4 に示す 6 区間に共通する特性として、区間の距離が長いこと、区間の途中で待ち時間の長い信号交差点が存在することが挙げられる。

機械学習の特徴量として加える、あるバス停間 1 区間の運行所要時間を順次入れ替え、予測精度の変化を調べた結果を、九州大学方面については表 4.5、表 4.6、表 4.7、九大学研都市駅方面については表 4.8、表 4.9、表 4.10 に示す。

これらの表から、運行所要時間の分散が最も大きい区間を加えた場合において、必ずしも予測精度の改善度が最大になるとは限らないとの結果が得られた。各経路・方面ごとに

表 4.5 指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（学園通經由九州大学方面）

区間始点	区間終点	MAE [秒]	MAAPE
(どの区間も加えない)		109.66	9.42%
九大学研都市駅	元岡郵便局前	95.02	7.84%
元岡郵便局前	元岡小学校前	107.00	8.92%
元岡小学校前	産学連携交流センター	101.99	8.25%
産学連携交流センター	西消防署元岡出張所前	102.89	8.42%
西消防署元岡出張所前	九大ビッグオレンジ	99.76	8.29%

表 4.6 指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（周船寺經由九州大学方面）

区間始点	区間終点	MAE [秒]	MAAPE
(どの区間も加えない)		136.28	6.90%
九大学研都市駅	北原	132.43	6.68%
北原	山崎	132.97	6.65%
山崎	周船寺東口	133.24	6.67%
周船寺東口	周船寺	141.25	7.03%
周船寺	泉	134.39	6.78%
泉	富士見	134.14	6.73%
富士見	たろう保育園前	136.40	6.84%
たろう保育園前	元岡農協前	135.38	6.82%
元岡農協前	九大東ゲート	135.81	6.69%

見ると次のような特徴があった。

九州大学方面について、学園通經由は、運行所要時間の分散が最も大きい区間の所要時間を特微量として加えたとき、予測精度が最も大きく改善した。その一方で、周船寺經由と横浜西經由は、分散が最大の区間の所要時間を特微量として加えたときの予測精度の改善度より、別の 1 区間を特微量として加えたときのほうが、予測精度をより改善させる効果があった。

九大学研都市駅方面について、学園通經由と横浜西經由は、運行所要時間の分散が最も大きい区間の所要時間を特微量として加えたとき、予測精度が最も大きく改善した。一方で、周船寺經由は、分散が最大となっていない別の 1 区間を特微量として加えたときのほうが、予測精度をより改善させる効果があった。

表 4.7 指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（横浜西経由九州大学方面）

区間始点	区間終点	MAE [秒]	MAAPE
(どの区間も加えない)		132.87	6.58%
九大学研都市駅	舞鶴高校入口	131.21	6.51%
舞鶴高校入口	横浜西	140.58	7.06%
横浜西	工芸会ワークセンター	134.16	6.64%
工芸会ワークセンター	玄洋高校	131.36	6.55%
玄洋高校	今出	128.48	6.30%
今出	元岡郵便局前	134.47	6.65%
元岡郵便局前	元岡小学校前	137.82	6.89%
元岡小学校前	九大東ゲート	127.45	6.19%

表 4.8 指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（学園通経由九大学研都市駅方面）

区間始点	区間終点	MAE [秒]	MAAPE
(どの区間も加えない)		88.76	8.52%
九大総合グラウンド	九大船舶・航空実験棟	87.85	8.25%
九大船舶・航空実験棟	九大農学部	87.64	8.41%
九大農学部	九大工学部	88.60	8.28%
九大工学部	九大ビッグオレンジ	87.69	8.24%
九大ビッグオレンジ	九大センターゾーン入口	87.35	8.20%
九大センターゾーン入口	西消防署元岡出張所前	87.93	8.20%
西消防署元岡出張所前	産学連携交流センター	85.80	8.04%
産学連携交流センター	元岡小学校前	85.52	7.91%
元岡小学校前	元岡郵便局前	86.84	8.22%
元岡郵便局前	九大学研都市駅	71.28	6.76%

ここで、機械学習器が予測するにあたって、どの特徴量が予測に大きな影響を及ぼしたかを、特徴量重要度の指標を用いて調べた。結果の上位 3 件を九州大学方面については表 4.11、表 4.12、表 4.13、九大学研都市駅方面については表 4.14、表 4.15、表 4.16、に示す。

重要度が高い値を示している区間の運行所要時間を特徴量として加えたときに、予測精度がよく改善していた。また、横浜西経由九州大学方面を除き、重要度が上位の特徴量と

表 4.9 指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（周船寺經由九大学研都市駅方面）

区間始点	区間終点	MAE [秒]	MAAPE
(どの区間も加えない)		114.67	6.79%
九大総合グラウンド	九大船舶・航空実験棟	110.29	6.54%
九大船舶・航空実験棟	九大農学部	116.12	6.88%
九大農学部	九大工学部	108.05	6.41%
九大工学部	九大ビッグオレンジ	109.80	6.45%
九大ビッグオレンジ	九大中央図書館	103.55	6.16%
九大中央図書館	九大イーストゾーン	117.47	6.95%
九大イーストゾーン	九大東ゲート	112.12	6.64%
九大東ゲート	元岡農協前	105.06	6.25%
元岡農協前	たろう保育園前	109.22	6.48%
たろう保育園前	富士見	112.00	6.64%
富士見	泉	99.87	5.96%
泉	周船寺	100.00	5.95%
周船寺	周船寺東口	107.64	6.39%
周船寺東口	山崎	105.68	6.28%
山崎	北原	113.99	6.75%
北原	九大学研都市駅	113.11	6.71%

して、運行時間のうち時の要素が出ていた。

4.6.2 路線の起点から路側機設置候補地点までのすべての区間それぞれの所要時間を加えた場合

4.3 節で述べたように、ある地点 A に路側機を設置したとき、地点 A では起点から地点 A までの各バス停間の運行所要時間すべてを得ることができる。そこで、各区間の運行所要時間を 1 区間ずつ入れ替えるのではなく、起点から指定の区間までのすべての区間の運行所要時間を特徴量として加えた場合の予測精度を求めた。結果を表 4.17, 表 4.18, 表 4.19, 九大学研都市駅方面については表 4.20, 表 4.21, 表 4.22 に示す。

九州大学方面、九大学研都市駅方面とも、特徴量として加える区間が増えるにつれて、全体的には予測精度が漸次改善していった。ただし、各表の最終行にあたる、目的地までの最終バス停区間の所要時間を得られた段階で、バスはすでに目的地に到着していること

表 4.10 指定したバス停間 1 区間の運行所要時間を加えた場合の予測精度（横浜西經由九大学研都市駅方面）

区間始点	区間終点	MAE [秒]	MAAPE
(どの区間も加えない)		92.06	6.35%
九大総合グラウンド	九大船舶・航空実験棟	87.04	6.03%
九大船舶・航空実験棟	九大農学部	86.21	5.98%
九大農学部	九大工学部	96.95	6.68%
九大工学部	九大ビッグオレンジ	86.02	5.96%
九大ビッグオレンジ	九大中央図書館	91.17	6.30%
九大中央図書館	九大イーストゾーン	85.55	5.95%
九大イーストゾーン	九大東ゲート	86.80	6.02%
九大東ゲート	元岡小学校前	84.24	5.85%
元岡小学校前	元岡郵便局前	86.03	5.99%
元岡郵便局前	今出	87.05	6.02%
今出	玄洋高校	85.62	5.94%
玄洋高校前	工芸会ワークセンター	90.03	6.22%
工芸会ワークセンター	横浜西	94.83	6.53%
横浜西	舞鶴高校入口	86.00	5.96%
舞鶴高校入口	九大学研都市駅	73.19	5.08%

表 4.11 特徴量重要度（学園通經由九州大学方面）

順位	特徴量	値	特徴量の意味
1	d_1to32	21.23	所要時間（九大学研都市駅→元岡郵便局前）
2	date_hour	17.10	運行日時（時）
3	d_42to51	15.11	所要時間（西消防署元岡出張所前→九大ビッグオレンジ）

から、この特徴量は予測に利用することができない。また、その手前までの区間の所要時間も、できる限り起点側に近いところまでだけの特徴量として利用して、所要時間の予測結果を早期に提示することが実用上は必要である。このことから、次の条件のもとに1時間前の運行所要時間の情報を利用しない場合の最適な路側機の設置地点を定めた。

全バス停間の運行所要時間を特徴量として加えたとき、その重要度が3位以内の区間

表 4.11～表 4.16 に示した特徴量重要度上位の特徴量のうち、バス停間の運行所要時間を表すものを選定する。さらにその中でも、より起点側に近い区間を採用

表 4.12 特徴量重要度（周船寺経由九州大学方面）

順位	特徴量	値	特徴量の意味
1	date_hour	8.93	運行日時（時）
2	d_6to8	7.98	所要時間（北原→山崎）
3	d_12to14	7.14	所要時間（周船寺→泉）

表 4.13 特徴量重要度（横浜西経由九州大学方面）

順位	特徴量	値	特徴量の意味
1	d_58to60	22.48	所要時間（九大船舶・航空実験棟→九大総合グラウンド）
2	d_28to30	18.47	所要時間（玄洋高校→今出）
3	d_54to56	9.15	所要時間（九大工学部→九大農学部）

表 4.14 特徴量重要度（学園通経由九大学研都市駅方面）

順位	特徴量	値	特徴量の意味
1	d_32to1	30.57	所要時間（元岡郵便局前→九大学研都市駅）
2	date_hour	18.29	運行日時（時）
3	d_40to34	9.35	所要時間（西消防署元岡出張所前→産学連携交流センター）

表 4.15 特徴量重要度（周船寺経由九大学研都市駅方面）

順位	特徴量	値	特徴量の意味
1	d_14to12	13.75	所要時間（泉→周船寺）
2	date_hour	12.68	運行日時（時）
3	d_8to6	11.72	所要時間（山崎→北原）

する。

別ルートに置いた路側機の位置が近い場合は、2つを統合できる位置

3ルートが分岐したあと、あるいは合流する前すぐの地点に設置するような場合は、コストを低減することを考えて、路線が分岐する手前や合流した後に設置する。

同一の道路上に近接した距離で置く場合は、比較してより起点側の位置

同じ道路上に各ルートで最適とされた設置位置が近接して定められた場合、起点側に近い地点に統合する。

これらの条件を満たすように設置地点の候補を定めると、

表 4.16 特徴量重要度（横浜西経由九大学研都市駅方面）

順位	特徴量	値	特徴量の意味
1	d_22to1	21.98	所要時間（舞鶴高校入口→九大学研都市駅）
2	date_hour	13.81	運行日時（時）
3	d_45to34	6.55	所要時間（九大東ゲート→元岡小学校前）

表 4.17 指定したバス停までの各区間所要時間を加えた場合の予測精度（学園通経由九州大学方面）

区間終点	MAE [秒]	MAAPE
（どの区間も加えない）	109.66	9.42%
元岡郵便局前	95.02	7.84%
元岡小学校前	93.38	7.76%
産学連携交流センター	88.07	7.10%
西消防署元岡出張所前	85.34	6.76%
九大ビッグオレンジ	81.12	6.35%

- 九州大学方面
 - － 元岡郵便局前バス停付近（学園通経由・横浜西経由）
 - － 山崎バス停付近（周船寺経由）
- 九大学研都市駅方面
 - － 産学連携交流センターバス停付近（学園通経由）
 - － 周船寺バス停付近（周船寺経由）
 - － 元岡小学校前バス停付近（横浜西経由）

となる。既設の路側機1台に対し、さらに追加で5台を必要とする結果となった。

4.7 解析 II: 1時間前の所要時間実績値を取得できる地点の選定

解析 I の結果は、運行所要時間の予測精度を路線全体で向上させる場合、路線全体の複数箇所に路側機を設置する必要があることを示している。解析 II では、特徴量「運行所要時間を予測する対象の区間における、1時間前の所要時間実績の平均値」を算出するために必要な路側機の設置地点を選定した。

表 4.18 指定したバス停までの各区分所要時間を加えた場合の予測精度（周船寺経由九州大学方面）

区間終点	MAE [秒]	MAAPE
(どの区間も加えない)	136.28	6.90%
北原	132.43	6.68%
山崎	131.55	6.55%
周船寺東口	126.21	6.23%
周船寺	126.74	6.23%
泉	123.40	6.06%
富士見	122.36	5.98%
たろう保育園前	118.23	5.78%
元岡農協前	118.61	5.77%
九大東ゲート	114.28	5.49%

表 4.19 指定したバス停までの各区分所要時間を加えた場合の予測精度（横浜西経由九州大学方面）

区間終点	MAE [秒]	MAAPE
(どの区間も加えない)	132.87	6.58%
舞鶴高校入口	131.21	6.51%
横浜西	134.55	6.70%
工芸会ワークセンター	132.49	6.60%
玄洋高校	133.34	6.57%
今出	125.35	6.07%
元岡郵便局前	127.09	6.19%
元岡小学校前	129.68	6.34%
九大東ゲート	120.29	5.67%

4.7.1 1 時間前の運行所要時間の取得状況

本研究における位置情報履歴データから運行履歴を算出するにあたり、本来バスが運行されている時間帯にもかかわらず、データ取得の失敗などにより当該時間帯の情報が得られない状況が見られた。このような場合、1 時間前の所要時間実績値を利用できなくなる。この状況の発生頻度を調査した。結果を表 4.23 および表 4.24 に示す。

表 4.20 指定したバス停までの各区間所要時間を加えた場合の予測精度（学園通經由九大学研都市駅方面）

区間終点	MAE [秒]	MAAPE
(どの区間も加えない)	88.76	8.52%
九大船舶・航空実験棟	87.85	8.25%
九大農学部	87.10	8.19%
九大工学部	91.00	8.39%
九大ビッグオレンジ	87.15	8.12%
九大センターゾーン入口	84.48	7.90%
西消防署元岡出張所前	84.12	7.81%
産学連携交流センター	80.30	7.51%
元岡小学校前	76.08	7.08%
元岡郵便局前	75.01	6.96%
九大学研都市駅	56.43	5.25%

各経路の便の全運行時間帯中、最大で約3分の1にあたる時間帯で、運行されているはずの便の位置情報履歴データが取得できていない状況が見られた。この原因として、路側機との通信不良で正しくデータが回収できなかった可能性や、位置情報履歴からの運行状況算出時に異常値として取り除かれた可能性が考えられる。

4.7.2 バス1運行における所要時間が1時間以内であるか

運行所要時間が取得できている時間帯であっても、その情報が予測対象便運行時間帯の1時間前までに取得できていなければ予測に利用することができない。本研究における対象路線・九州大学線は、時刻表上での最長所要時間が30分となっている。また、どの経路の路線であっても、路側機が1機すでに設置されている九州大学伊都キャンパス構内入口のほか、九大学研都市駅付近を必ず通行する。このことから、増設を1箇所のみとする場合は、九大学研都市駅付近に設置することが有効であると考えられる。実際に有効であるかどうか、位置情報履歴データから算出した運行状況を解析した。

既設の路側機は、九州大学伊都キャンパス構内への入構ゲート部に設置されている。バス停の区間との対応は、九州大学方面の九大ビッグオレンジ→九大理学部間となる。したがって、この区間を通過したバスが、1時間以内に当該路側機か、九大学研都市駅付近を通行すればよい。大学到着後、すぐに折り返さず大学構内のバス待機場場に駐車されるか、近接するバスの運行拠点・昭和バス伊都営業所へ回送される場合があるが、この場合は位

表 4.21 指定したバス停までの各区分所要時間を加えた場合の予測精度（周船寺経由九大学研都市駅方面）

区間終点	MAE [秒]	MAAPE
(どの区間も加えない)	114.67	6.79%
九大船舶・航空実験棟	110.29	6.54%
九大農学部	99.04	5.91%
九大工学部	114.12	6.77%
九大ビッグオレンジ	107.12	6.36%
九大中央図書館	94.25	5.64%
九大イーストゾーン	98.19	5.86%
九大東ゲート	99.85	5.94%
元岡農協前	94.21	5.63%
たろう保育園前	94.74	5.65%
富士見	93.26	5.57%
泉	91.28	5.47%
周船寺	87.64	5.24%
周船寺東口	79.15	4.78%
山崎	80.67	4.84%
北原	75.97	4.56%
九大学研都市駅	68.83	4.14%

置情報履歴の蓄積が路側機の設置地点から営業所までの短区間に限られるため、1 時間以内にその情報が取得できなくとも問題はない。

このことから、実際の解析にあたっては、大学を出発したバスが 1 時間以内に九大学研都市駅に到達するか、また、逆に九大学研都市駅を出発したバスが 1 時間以内に大学に到達するか、バスが運行するすべての日・時間・路線について解析した。

位置情報履歴データから算出した 2019 年度 1 年間の運行履歴について、路線の起点から終点までの運行所要時間の統計値を求めたところ、所要時間が 1 時間を超えていると算出された例が数件見られた。それらはバス停間の所要時間に異常値が見られたことからエラーとして除去した。除去後の統計値を表 4.25 および表 4.26 に示す。その結果、すべての運行履歴について、1 時間以内に起点から終点まで到達していることが明らかとなった。

表 4.22 指定したバス停までの各区間所要時間を加えた場合の予測精度（横浜西経由九大学研都市駅方面）

区間終点	MAE [秒]	MAAPE
(どの区間も加えない)	92.06	6.35%
九大船舶・航空実験棟	87.04	6.03%
九大農学部	86.26	5.97%
九大工学部	84.06	5.84%
九大ビッグオレンジ	84.81	5.88%
九大中央図書館	85.78	5.94%
九大イーストゾーン	87.09	6.03%
九大東ゲート	82.76	5.74%
元岡小学校前	82.62	5.74%
元岡郵便局前	80.39	5.59%
今出	76.64	5.35%
玄洋高校	76.76	5.36%
工芸会ワークセンター	77.41	5.40%
横浜西	76.07	5.30%
舞鶴高校入口	76.35	5.32%
九大学研都市駅	60.78	4.23%

表 4.23 2019 年度運行所要時間取得状況（九大学研都市駅→九州大学）

経由	運行時間帯数	取得成功時間帯数	成功時間帯数の割合
学園通	6,099	4,965	81.4%
周船寺	5,976	4,037	67.6%
横浜西	6,342	4,643	73.2%
総計	18,417	13,645	74.1%

4.7.3 路側機が1台である場合の、九大学研都市駅方面の便における1時間前の運行所要時間の取得可否

路側機が大学構内入口に1台だけ設置されている現在の状況では、九大学研都市駅発九州大学方面の便は、大学到着直後に位置情報履歴を収集することができ、ただちに運行所要時間を算出することができる。一方、九州大学発九大学研都市駅方面の便は、大学を出

表 4.24 2019 年度運行所要時間取得状況（九州大学→九大学研都市駅）

経由	運行時間帯数	取得成功時間帯数	成功時間帯数の割合
学園通	5,976	4,939	82.6%
周船寺	6,096	4,392	72.0%
横浜西	5,976	4,036	67.5%
総計	18,048	13,367	74.1%

表 4.25 2019 年度運行所要時間統計値（九大学研都市駅→九州大学・異常値除去後）（秒）

経由	最小値	平均値	最大値	分散 σ^2
学園通	780	1,024.19	2,025	16,521.27
周船寺	1,093	1,605.22	2,630	23,534.04
横浜西	981	1,384.75	2,388	14,527.32

表 4.26 2019 年度運行所要時間統計値（九州大学→九大学研都市駅・異常値除去後）（秒）

経由	最小値	平均値	最大値	分散 σ^2
学園通	589	1,018.77	2,546	20,235.12
周船寺	1,216	1,584.23	2,651	17,025.69
横浜西	968	1,366.00	2,158	12,122.93

発した便が 1 時間以内に戻ってきて再度路側機を通過しない限り、予測対象便の 1 時間前の運行所要時間を算出することができない。

そのため、次の 2 段階の手順により、大学発九大学研都市駅方面について、1 時間前の運行所要時間実績値の取得が可能であるかどうかを確かめた。まず、各経由で位置情報履歴を収集できず、運行履歴を求めることができなかった日時を調査した。

2022 年 4 月 10 日の 15 時台を例にとり、学園通経由の便で 1 時間前の運行所要時間が取得できているかの確認手法を説明する。

1. 4 月 10 日の 14 時台に、九州大学→九大学研都市駅間を学園通経由で走った便があるかどうかを調べる。
2. 上記の便で走った車両のうち、少なくとも 1 台が、15 時までに路側機がある大学構内入口の路側機があるバス停間を通過しているかを確かめる。

この手順で便が検出されなかった場合、当該時間帯の運行所要時間予測に、1 時間前の運行所要時間を用いることができなくなる。この手順により、バスの運行がある各日付

表 4.27 1時間前の運行所要時間を取得できた数（学園通経由・九州大学→九大学研都市駅）

運行時間帯の総数	1時間以内に戻ってきた時間帯数	割合
9,775	8,352	85.4%

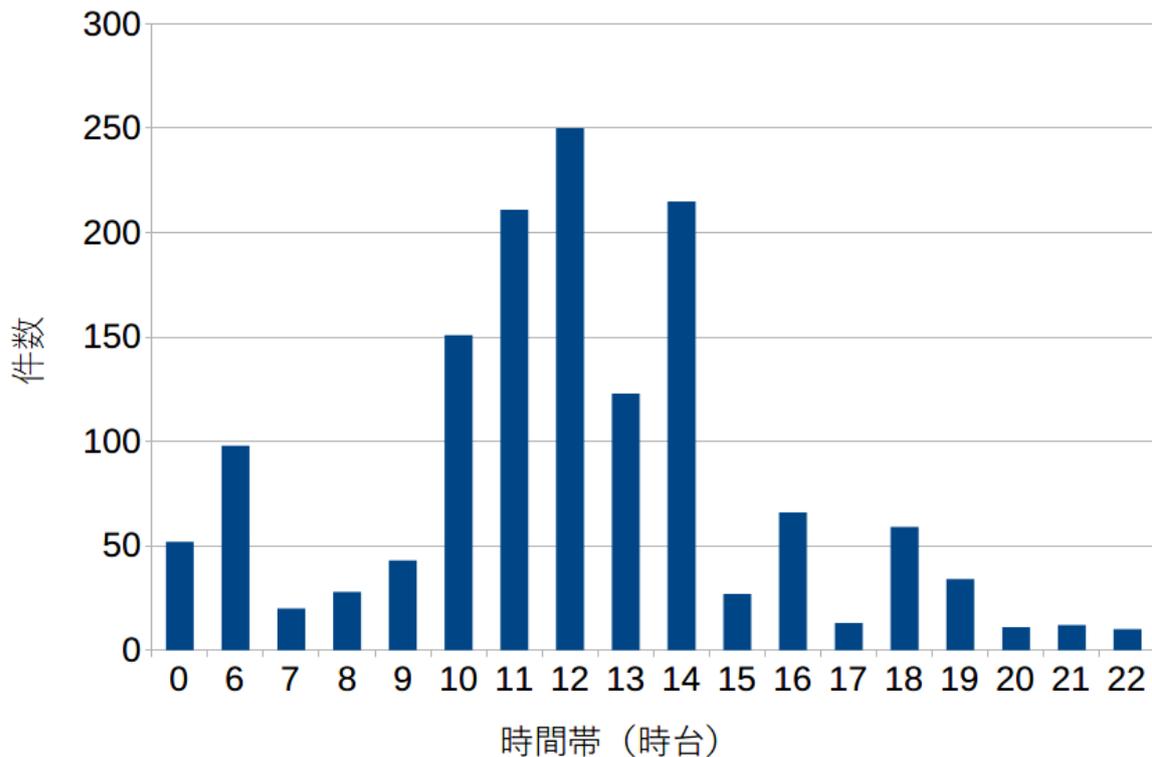


図 4.1 1時間以内に当該時間帯の運行所要時間情報が得られなかった件数（学園通経由・九州大学→九大学研都市駅）

の各時間帯に対して、1時間前の所要時間を算出できるかどうかを調べた。解析結果を表 4.27 と図 4.1 に示す。ここで、バス運行時間帯の総数とは、学園通経由の便が運行されていると検出できた時間帯の数（1時間区切り、たとえば 15 時台に運行があれば 1 件と数える）を示す。また、1時間以内に戻ってきた時間帯数は、当該時間帯に 1時間以内に路側機がある大学入口まで戻ってきた車両が存在していた時間帯の数である。

表 4.27 を見ると、全運行日時（日付および時間帯）に対して、およそ 15% の時間帯数で、当該時間帯の運行所要時間情報が 1時間以内に得られなかった。その件数を図 4.1 で確認すると、10～14 時の昼間時間帯に多くなっていた。

4.8 考察

4.8.1 予測精度をよく改善する区間と運行所要時間の分散との関係

予測精度を改善するために、機械学習器に加える特徴量として、どのバス停区間の運行所要時間情報を用いるとよいかを選定する指標として、各バス停区間の運行所要時間の分散を求めた。運行所要時間の分散が大きい区間は、運行便によって運行所要時間の差が大きいことを意味する。運行所要時間の差が大きいことはバスの遅延に関係が深いと考えた。しかし、実際には、予測精度をよく改善する区間は、必ずしも運行所要時間の分散が大きい区間とは限らなかった。

4.8.2 予測精度をよく改善する区間と特徴量重要度との関係

どの特徴量が予測精度の向上に寄与していたかを、特徴量重要度の値によって比較した結果、特徴量重要度の値が大きい区間は、予測精度の改善に大きく貢献していることがわかった。しかし、ルートによっては、各特徴量重要度の値に大きな差が見られない状況もあった。また、各区間の運行所要時間以外の特徴量として、運行時間帯の情報が出現していた。

4.8.3 1時間前の起終点間の運行所要時間の取得可能性

表 4.27 では、路側機が大学入口の 1 箇所だけにある場合に、路側機設置箇所を 1 時間以内にバス車両が再び通過するかどうかをデータから確認した。その結果、全体の 15% の時間帯において、1 時間以内にバス車両が再度通過せず、次時間帯の予測の際に実際には予測対象時間帯の 1 時間前の運行所要時間情報が利用できないことが明らかになった。本論文では、予測対象便の 1 時間前の運行所要時間を利用して予測したが、もしこれを 30 分前や 15 分前など、より予測対象便の運行時間に近い情報を利用したい場合には、路側機 1 箇所だけではデータ取得が間に合わないことを示している。

また、1 時間以内にバス車両が路側機を再度通過しなかった時間帯として、10～14 時の昼間時間帯が多く検出された。これは、当該時間帯のバス運行本数が比較的少なく、路側機を通過した後、路線の端点である九大学研都市駅、大学構内、および大学に隣接するバス営業所で待機するために、路側機を再度通過するまでに 1 時間以上を要しているものとみられる。

4.8.4 路側機増設箇所

前章の表 3.4 および表 3.5 から、予測対象便が大学内やある一定の地点に到達するまでの運行所要時間を算出して起終点間の所要時間を予測するよりも、予測対象便の1時間前の起終点間の運行所要時間を用いて予測する方が、予測精度は大幅に高まることが分かった。起終点間の所要時間を迅速に取得するためには、起点と終点に路側機があることが望ましい。現在、路側機は路線の一方の端点に近い場所である九州大学構内入口に1箇所設置されている。

本研究で対象とした九州大学線は、調査対象期間中に最も時間がかかるルートで最も遅れた場合でも1時間を要しなかったことが確認できている。1時間前の運行所要時間を取得利用するためには、九大学研都市駅付近に路側機を追加設置することが有効であると考えられる。

しかし、4.7.1, 4.7.2, 4.7.3の結果から、予測対象時間帯の1時間前の運行所要時間平均値を必ずしも取得できる状況ではないと分かった。そのため、1時間前の運行所要時間を取得できない時間帯を予測する場合は、九大学研都市駅付近以外に4.6節で候補として挙げた5箇所の設置地点のいずれかに設置して、途中バス停までの所要時間情報を加えることにより、予測精度を向上させる手法を併用すると精度を維持できると推測される。

4.9 まとめ

本章では、バスの運行所要時間の予測精度を高めるために位置情報履歴データを収集する路側機を追加設置するとき、その設置コストと、当該路側機で取得できる情報を追加して運行所要時間を予測する場合の精度の変化を解析し、適切な設置台数と設置位置を検討した。

その結果、最適な設置位置は必ずしもバス停間運行所要時間の分散の最も大きな箇所と同じにはならないことを確認した。また、機械学習によるシミュレーションを実施し、予測精度を最も向上させる区間は、特徴量重要度の値も大きいことを確認した。次いで、リアルタイムな情報を得られない場合を解決する際に有効な特徴量である「1時間前における予測対象区間の運行所要時間」を収集できる地点を検討し、路側機は路線の起点と終点に各1台ずつ設置し、1時間前の起終点間の運行所要時間を迅速に収集することが適当であるとした。一方で、その情報が得られない時間帯が、全運行時間帯数のおよそ4分の1存在することが明らかになった。そうした時間帯でも予測精度を維持するためには、路線の途中区間までの所要時間を収集して特徴量として用いることが有力であり、その際の選定基準は次が適当であると結論付けた。

-
- 全バス停間の運行所要時間を特徴量として加えたとき，その重要度が3位以内の区間，かつ起点側に近いもの。
 - 別ルートに置いた路側機の位置が近い場合は，2つを統合できる位置。
 - 同一の道路上に近接した距離で置く場合は，比較してより起点側の位置。

第5章

急ブレーキ事象の識別と要因解明

5.1 はじめに

位置情報履歴データの活用方法に、交通事故状況の分析と予防がある。自動車の走行挙動に関するデータのうち、事故や事故のおそれとなる事象に密接に関係するものが、車両に生じる大きな加速度である。大きな加速度が発生した時は、当該車両に急加速や急ブレーキ、急ハンドルといった挙動が発生している可能性がある。これらの挙動は危険回避のために行われる他、場合によっては事故に直接かかわっていることもある。

本章では、車両の危険な走行挙動のうち、後方向への加速度としてデータに現れる急ブレーキに着目する。急ブレーキが発生する要因は、前方を走行する車両の急ブレーキへの対応、車両や人の飛び出しに対する事故回避、車両のスリップなどが挙げられる。事故情報やヒヤリハット事象を収集分析すると、急ブレーキの直接的要因を詳細に捉えることが可能だが、それらの直接的要因の陰に、急ブレーキを発生させるもととなる事故誘発要因が隠れている可能性がある。従来は、危険な走行挙動を車両の速度や加速度の値を用いて定義し、定義に合う状況を数え上げる頻度解析を実施するなど、統計的手法に基づくアプローチが行われてきた。解析の結果、急ブレーキが発生する地点を明らかにしている研究は数多い [53, 54]。しかし、その地点で急ブレーキを発生させる真の要因を探るには、各地点の状況を人間の手によって観察し、検討する必要があった。近年では、機械学習を利用してデータを解析する試みもさかんである。機械学習を利用すると、急ブレーキの真の要因を、急ブレーキの事象を抽出するのに寄与した特徴量として明らかにできる可能性がある。そこで、本章では、位置情報履歴データから取り出せる情報や、その組み合わせから特徴量を算出し、機械学習によって急ブレーキが発生しているか否かの識別を実施した。識別に有効であった特徴量を求め、その特徴量や組み合わせの中に、急ブレーキ発生要因を示す特徴量やその組み合わせが存在しているかを調査した。

その結果、急ブレーキの識別に寄与した特徴量として、年・月や曜日といった、時間的

要因を示すものが抽出された。特徴量のラベルが表している日時における位置情報を地図上にプロットしたところ、急ブレーキ発生頻度の上位に抽出された地点とは別の地点が抽出された。

5.2 データセット

本章では、データセットとして会津若松市公用車・公共交通車両走行情報を利用した。データセットの概要は2.4節の通りである。

5.3 危険な走行挙動「急ブレーキ」

本章では、自動車の危険な走行挙動として「急ブレーキ」を取り上げる。急ブレーキは、運転者が道路上の障害物や急に飛び出して来た人、急ブレーキや車線逸脱などが発生した車などとの衝突を防止する際に発生する。いずれも、事故発生や、事故発生のリスクを示す重要な事象である。急ブレーキが発生するか否かを機械学習によって識別し、識別に寄与した特徴量を調べると、急ブレーキの条件を発見できると考えられる。急ブレーキは、車の進行方向と逆方向の加速度が発生した状況としてデータセットに出現する。本章で利用するデータセットにおいては、y軸方向の加速度として記録される。急ブレーキを判別する際の加速度閾値は、 $-0.25G$ とする研究と、 $-0.5G$ とする研究の双方が存在する。本章では、閾値を $-0.25G$ と定めて解析を実施した。

5.4 急ブレーキ発生状況の基礎解析

本章における解析にあたり、急ブレーキの発生が検出された地点における、急ブレーキの発生頻度と、その地点における走行記録の総数との比較を実施した。同じ地点の走行記録とみなすために、一定の範囲の位置情報を統合した。具体的には、緯度・経度の十進法度数表記のうち小数第4位以下を四捨五入して利用した。この処理により、会津若松市付近における緯線方向約110.9m、経線方向約88.6mの四角形の領域内の点群の座標が同一座標にあるものと見なされる。急ブレーキ発生地点と、その地点における走行履歴数および急ブレーキ記録件数、急ブレーキの割合を表5.1に示す。

ここで、急ブレーキ件数とは、データに記録されている前後方向加速度が $-0.25G$ 以下となっているものを単に数えたものとなる。従って、一連の運転における1回のブレーキであっても、その加速度が検出され続けている限り、件数としては複数カウントされる。

その結果、急ブレーキ発生頻度が上位でありながら、その地点の走行記録全体との比率では急ブレーキの確率が低い地点、逆に、走行記録数が少ないながら、その多くが急ブ

表 5.1 急ブレーキ発生地点と急ブレーキの割合

急ブレーキ件数	走行履歴数	発生割合	地点を表す特徴量名
772,586	778,557	0.9923	p:37.549x139.933
289,707	555,948	0.5211	p:37.506x139.929
106,608	132,199	0.8064	p:37.549x139.932
42,663	89,328	0.4776	p:37.490x139.933
31,399	75,915	0.4136	p:37.507x139.931
22,340	49,321	0.4530	p:37.500x139.892
22,044	56,248	0.3919	p:37.490x139.932
20,103	49,759	0.4040	p:37.507x139.932
19,777	40,708	0.4858	p:37.492x139.919
19,559	22,501	0.8693	p:37.489x139.933
16,777	35,308	0.4752	p:37.498x139.921
16,385	35,296	0.4642	p:37.491x139.922
15,751	18,419	0.8551	p:37.548x139.932
14,703	33,609	0.4375	p:37.500x139.934
14,156	33,785	0.4190	p:37.499x139.925
14,023	33,998	0.4125	p:37.490x139.930
13,530	30,625	0.4418	p:37.492x139.918
13,254	13,324	0.9947	p:37.548x139.933
13,139	29,077	0.4519	p:37.498x139.924
12,724	28,807	0.4417	p:37.501x139.920

レーキとして記録されている地点が見られた。たとえば、表 5.1 中で発生割合の値を太字で示した点は、走行履歴数に対する急ブレーキ発生件数の割合が 50% を下回っている。一方、点 “p: 37.548x139.933” や点 “p:37.548x139.932” のように、走行履歴のほとんどが急ブレーキと判定されているような場所も存在していた。急ブレーキの割合が極めて高くなっている地点として、表 5.1 の中から、発生割合が 80% 以上になっていた 4 地点について個別に調査すると、車両の運行拠点や、信号交差点となっていた。これらの場所では、車両駐車時の停止直前や、交差点停止時のブレーキ踏み込みにより、 $-0.25G$ 以下の加速度が記録されたものと考えられる。

ここで、急ブレーキ発生件数の上位 50 地点を地図上にプロットしたものを図 5.1 に示す。

図 5.1 からは、急ブレーキ件数が多かったのは主に会津若松市の中心部であったことが



図 5.1 急ブレーキ発生件数の上位 50 地点

わかる。一方、特定された場所は、急ブレーキが検出された時刻がさまざまであったため、地図からそれ以上の状況を推測することが不可能であった。このことから、運転手が急ブレーキを使用する条件に、場所以外の要因がかかっている可能性が高いとの仮説を立てた。この仮説を検証するため、機械学習を利用して、位置情報履歴データから急ブレーキの事象を識別するモデルを構築し、急ブレーキ識別に大きく作用する特徴量を検証した。

5.5 機械学習手法を用いた急ブレーキの識別

5.5.1 手法概略

本章では、機械学習手法として、サポートベクターマシン (SVM: Support Vector Machine) と特徴選択 (FS: Feature Selection) を組み合わせた SVM+FS 手法を利用した。SVM は二値分類器で、入力データを正例と負例に分類する。本章では、急ブレーキを意味する正例として、約 2.6 億件のデータセットから y 軸方向（前後方向）加速度

表 5.2 SVM+FS モデル構築に利用したデータ数

区分	データ数
正例	2,904,504
負例	2,905,667
計	5,810,171

表 5.3 選択した特徴量

ID	記号	説明
1	k	車両 ID
2	y	年 (2015 – 2018)
3	m	月 (1 – 12)
4	w	曜日 (mon, ..., sun)
5	t	時 (0 – 23)
6	p	地点 (緯度・経度)
7	x	x 軸方向 (左右方向) 加速度
8	y	y 軸方向 (前後方向) 加速度
9	z	z 軸方向 (上下方向) 加速度
10	d	進行方向方位 (1 – 12)
11	v	速度

が $-0.25G$ 以下のデータを抽出した。正例のデータ総数は 2,904,504 件であった。正例以外がすべて急ブレーキでない負例となるが、元のデータセットの規模が大きいため、負例のデータ数が正例のデータ数と比べて極めて多くなる。正例と負例の数が不均衡の状態のまま SVM モデルを構築すると、適切な識別ができない。そのため、負例として正例とほぼ同数のデータを負例としてランダムに取り出した。その件数を表 5.2 に示す。

本章では、急ブレーキに関連すると考えられる特徴量を網羅的に探索し、表 5.3 に示す 11 個の特徴量を選び出した。詳細を以下に示す。

5.5.2 日時

車両の走行挙動は季節、曜日、時間帯により異なると考えられる。本章では、これらの情報を説明変数として採用した。位置情報データ中の走行日時フィールドから年・月・曜日・時を抽出し、それぞれ記号 y , m , w , t と表現した。この記号と実際の値を結合し、特徴量を生成した。たとえば、「2018 年 5 月 1 日 8 時 15 分」というデータからは、“ y :

2018”, “m:5”, “w:tue”, “t:8” という特徴量が生成される。

5.5.3 車両走行位置

車両走行位置は緯度と経度の組で表される。各値の小数第4位で四捨五入し、記号 p を付与して特徴量を生成した。たとえば、地点 (37.524830,139.937097) は、特徴量 “p:37.525x139.937” と表される。

また、車両挙動は車両の進行方向方位によって異なる傾向にある。たとえば、同じ道路の同じ区間にある坂道を通過する際、下り坂の方向となる進行方向方位では急ブレーキの挙動が増え、上り坂の方向となる進行方向方位では減る。従って、本章では車両の進行方向方位を急ブレーキの要因の候補として調査した。車両の進行方向方位は、2点の位置情報を利用し、その時系列による移動方向を採用した。方位を表現するにあたり、詳細な方位角は急ブレーキの発生箇所の表現には必要ないと考えたため、本章では方位を時計に見立て、北を12時、東を3時、南を6時、西を9時として12方位で表現し、記号 d を付与して特徴量とした。たとえば南方向に走行している場合は、特徴量 “d:6” と表される。

5.5.4 特徴量の組み合わせによる新たな特徴量の生成

5.5.3 までに示した特徴量は、表 5.3 に示した特徴量のうち、7番、9番、11番を除く8つである。これらの組み合わせのうち、どの特徴量も選択しない例を除くと、 $2^8 - 1 = 255$ 通りの組み合わせ数となる。さらに、各特徴量をその記号とともに結合したものを、新たな特徴量として追加した。たとえば、結合により “kymwt:20:2016:08:mon:13” という特徴量が生成される。これは、「車両 ID が 20, 2016 年, 8 月, 月曜日, 13 時」を意味する。このようにして生成した特徴量について、5 件以上出現したものを数えると、182,779 個の特徴量となった。これらのうち、走行位置に関する特徴量は 9,730 個、それ以外の車両 ID や走行日時、車両の進行方向方位の特徴量は 173,049 個であった。

5.5.5 SVM+FS 手法

SVM は膨大な特徴量を持つデータを二値分類する際に良い性能を示す分類器のひとつである [55]。一方で、あまりにも特徴量が多すぎると SVM の識別性能が低下し、また、少数の特徴量を利用して構築した SVM 分類器が、全特徴量を利用して構築した SVM 分類器と遜色のない識別性能を有するという報告もある [56]。特徴量が多くなると計算量が増大するため、同等の識別性能を得られるのであれば、特徴量を減らして学習・推論することが、計算の高速化に有効である。SVM+FS 手法では、各データについて特徴量をカテゴリ変数化して one-hot 形式で表現し、値が 1 となった特徴量ラベルを「特徴語」とし

て抽出して、特徴語の群をひとつの文書 d_i とする。この文書に対して特徴選択の指標となる svm-score を算出し、特徴選択を実施した。アルゴリズムを次に示す。

svm-score の計算

1. N 個の文書集合 D を M 個のクラスに分類する。本章の場合は $M = 2$ 。
2. 文書 $d_i \in D (1 \leq i \leq n)$ が、 m 個の相異なる単語 $w_{i,j} (1 \leq j \leq m)$ を含むとき、文書 d_i から、単語を 1 つだけ含むように m 個の文書 $d_{i,j}$ を生成する。
3. 各文書 $d_{i,j}$ を単語ベクトル $v(d_{i,j}) = v_1, \dots, v_k, \dots, v_n$ に変換する。ここで、 n は、文書集合 D に含まれる相異なる単語の総数である。 $w_{i,j}$ が $v(d_{i,j})$ 中の k 番目の要素に対応する場合に $v_k = 1$ とし、そうでない場合は $v_k = 0$ とする。
4. 各分類クラスについて、文書 d_i が当該クラスに属するとき、各文書 $d_{i,j}$ に対して正例のフラグを割り当てる。そうでない場合は負例のフラグを割り当てる。
5. ステップ (3) で生成された単語ベクトルの集合をもとに、SVM による文書分類モデルを構築し、各文書 $d_{i,j}$ の重みを求める。ここで、文書 $d_{i,j}$ はただ 1 つの単語を含むため、文書の重みは各単語 $w_{i,j}$ の重みに等しい。この重みを svm-score と呼ぶ。

この svm-score を利用して、SVM+FS は次の 3 ステップで実行される。

SVM+FS

1. svm-score が上位 K 個の正例の語と、上位 K 個の負例の語を選択する。
2. ステップ (1) で選択した各 K 個の語からなる入力ベクトルに、各文書 $d_{i,j}$ を変換する。
 - この入力ベクトルは、 K 個の正例の語と K 個の負例の語からなる hot-BoW ベクトルである。つまり、文書中に 1 つ以上含まれる語に対応する要素の値が 1 となり、それ以外の要素は 0 となる。
 - もし文書が分類対象クラスに属している場合は、その文書から変換された入力ベクトルには正例のフラグが割り当てられ、そうでない場合は負例のフラグが割り当てられる。
3. ステップ (2) で生成された入力ベクトルを SVM を用いて学習し、分類モデルを構築する。

5.6 実験

5.6.1 SVM+FS によるモデル構築

表 5.2 に示した約 580 万件のデータを学習用データとして SVM モデルを構築し、各特徴量の svm-score を算出した。5.5.4 節で述べた通り、特徴量の総数は 182,779 個である。SVM のアルゴリズムには SVM^{perf} [57] を利用した。 SVM^{perf} は線形カーネルを利用しており、高速な計算が可能である。アルゴリズムのパラメータとしては、 $C = 20$ としたほかはデフォルト値を採用した。

5.6.2 評価指標

識別性能の評価には、Accuracy, Precision, Recall, F 値の 4 つを採用した。各指標の定義は 2.6.1 節に示している。

5.6.3 急ブレーキにおける地点の要因の影響評価

急ブレーキの発生に影響する要因のうち、地点の要素の影響を評価した結果を表 5.4 と表 5.5 に示す。これらの表は、急ブレーキを識別する際の地点に関する要因の効果を示したものである。各 1 行目の $K = ALL$ は、182,779 個すべての特徴量を利用した場合の結果であり、2 行目以降は上位各 K 個の正例の語、および負例の語を利用した場合の結果である。各 K の場合について 5 分割交差検証を実施した。この結果、急ブレーキの識別にあたって最良の F 値を示したのは、学習用データに対しては $K = 3000$ のとき 71.5%、全データに対しては $K \geq 50$ のとき 57.7% であった。この結果は、いずれのデータ群でも特徴選択がうまく機能せず、地点の特徴を用いた急ブレーキの識別性能が高くないことを意味している。図 5.2 は、各地点における急ブレーキの発生頻度と発生確率の関係を示したものである。図 5.2 の通り、急ブレーキが頻繁に発生する位置では、必ずしも急ブレーキの発生確率が高いわけではないことが分かる。したがって、地点を表現する特徴量は必ずしも急ブレーキに影響を与えるわけではないと言える。

5.6.4 地点の要因以外の特徴量の効果

地点を表現する特徴量が必ずしも急ブレーキに影響を与えるわけではないことから、地点の特徴量以外を利用した SVM モデルを構築した。svm-score が上位となった特徴量のうち、正例の上位 10 件と負例の上位 10 件を表 5.6 と表 5.7 に示す。

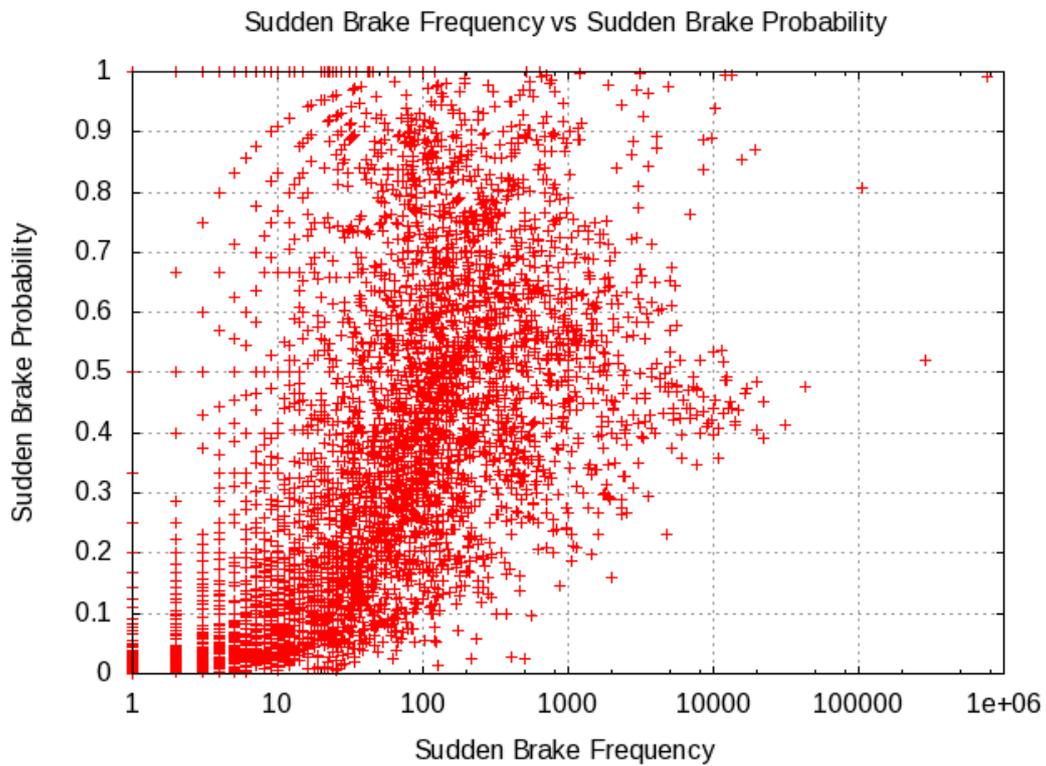


図 5.2 急ブレーキの発生確率と発生件数の関係

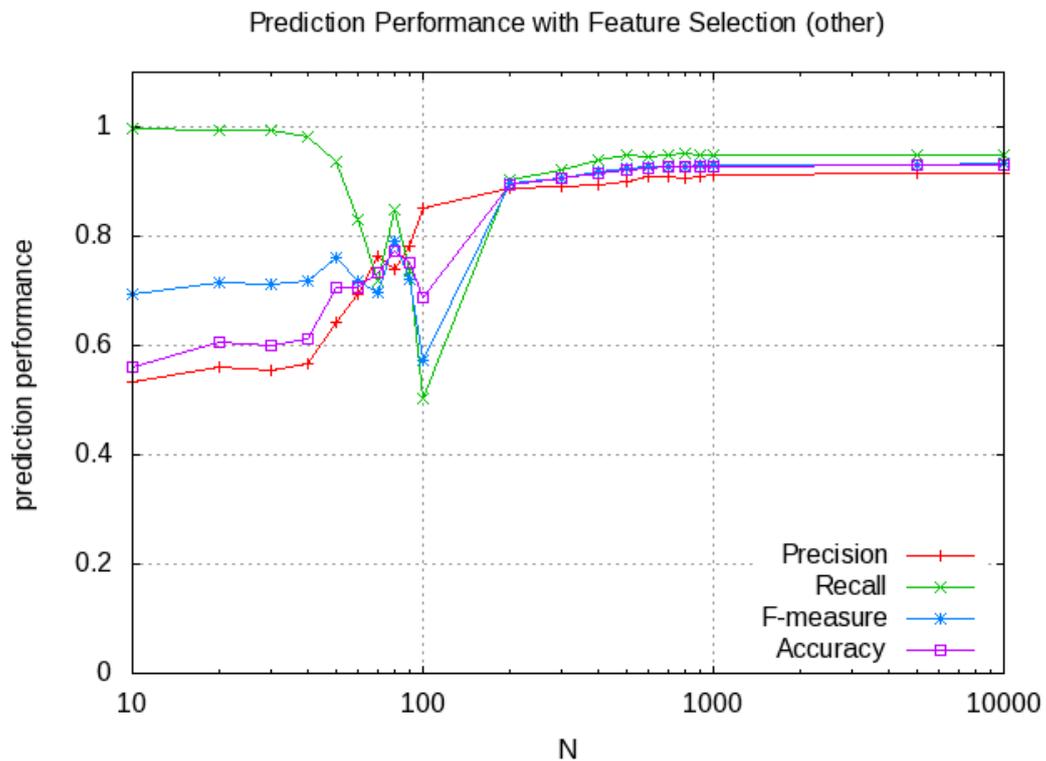


図 5.3 特徴選択を行った場合の急ブレーキ識別性能（地点の特徴量を除いたとき）

表 5.4 急ブレーキに影響する地点の要因の効果 (学習用データ)

K	Precision	Recall	F 値	Accuracy
ALL	0.7229	0.6803	0.7008	0.7097
10	0.5028	0.9989	0.6689	0.5057
50	0.5071	0.9908	0.6708	0.5140
100	0.5139	0.9864	0.6758	0.5268
500	0.5336	0.9611	0.6863	0.5607
1,000	0.5456	0.9547	0.6943	0.5798
2,000	0.5694	0.9298	0.7063	0.6134
3,000	0.6119	0.8602	0.7146	0.6566
4,000	0.6727	0.7463	0.7070	0.6910
5,000	0.7181	0.6850	0.7012	0.7081
6,000	0.7181	0.6850	0.7012	0.7081
7,000	0.7181	0.6850	0.7012	0.7081

図 5.3 と表 5.8 は、本実験で構築した急ブレーキ識別モデルを学習用データに適用した場合の識別性能を示したものである。同様に、モデルを全データに対して適用した場合の識別性能を図 5.4 および表 5.9 に示す。どちらの場合も識別性能はほぼ同等で、平均の F 値は約 93.2% を示した。このモデルは、地点の情報のみを利用して構築した急ブレーキ識別モデルよりも識別性能が良い。したがって、地点の情報以外の要因に関連する事象が、運転者の急ブレーキ動作に影響するという仮説が証明されたと言える。例として、正例の特徴量の第 4 位のラベル “kym:10: 2017:02” にあてはまる位置情報履歴のうち、上位 50 件を地図にプロットしたものを図 5.5 に示す。この特徴量は、2017 年 2 月に国道 294 号の猪苗代湖西方にある直線区間を走行していた車両 ID 10 の位置情報履歴データで検出された急ブレーキ事象を示している。この事象を示す特徴量は車両 ID・走行日時・年の情報・走行日時の月の情報であり、この急ブレーキ事象の識別には地点の情報以外の情報が寄与していたことを示唆すると考えられる。

5.7 まとめ

本章では、会津若松市公用車・公共交通車両走行情報オープンデータセットに対して、サポートベクターマシンを利用した機械学習と、特徴選択を組み合わせた手法により、自動車の危険な走行挙動のひとつである「急ブレーキ」の発生条件抽出を試みた。また、急ブレーキに関連する特徴を検出し、その特徴に含まれる位置にマッピングするマイニング

表 5.5 急ブレーキに影響する地点の要因の効果（全データ）

K	Precision	Recall	F 値	Accuracy
ALL	0.8068	0.4495	0.5774	0.6710
10	0.8038	0.4320	0.5620	0.6634
50	0.8068	0.4495	0.5774	0.6710
100	0.8068	0.4495	0.5774	0.6710
500	0.8068	0.4495	0.5774	0.6710
1,000	0.8068	0.4495	0.5774	0.6710
2,000	0.8068	0.4495	0.5774	0.6710
3,000	0.8068	0.4495	0.5774	0.6710
4,000	0.8068	0.4495	0.5774	0.6710
5,000	0.8068	0.4495	0.5774	0.6710
6,000	0.8068	0.4495	0.5774	0.6710
7,000	0.8068	0.4495	0.5774	0.6710

表 5.6 svm-score が上位の正例の語

順位	svm-score	急ブレーキ件数	全履歴数	特徴語
1	0.3753	20,218	23,696	kymwt:20:2017:08:tue:08
2	0.3534	40,244	42,525	kymwt:20:2017:08:wed:13
3	0.3432	20,218	23,789	kmwt:20:08:tue:08
4	0.3225	55,136	62,230	kym:10:2017:02
5	0.3016	34,097	34,137	ymwt:2016:07:tue:15
6	0.2961	20,303	26,529	ymwt:2017:08:tue:08
7	0.2904	5,667	7,114	kymwt:20:2016:08:mon:13
8	0.2825	2,543	3,132	kymwt:3:2017:08:wed:15
9	0.2804	84,670	113,663	kmwt:0:01:fri:08
10	0.2804	84,670	113,663	ymwt:2016:01:fri:08

と可視化ツールを発表した。可視化処理により、地図上に配置された特徴は、場所を示すだけでなく、急ブレーキを引き起こす新たな要因を見つける手がかりとして利用することができる。

解析にあたっては位置情報データセットを網羅的に探索し、特徴量の ID と特徴量の値の組を組み合わせて多数の特徴量を算出し、その特徴量に対して SVM+FS 手法を適用した。その結果、F 値の平均値で 93.2% のと、良い精度で急ブレーキの挙動を識別すること

表 5.7 svm-score が上位の負例の語

順位	svm-score	急ブレーキ件数	全履歴数	特徴語
1	-0.2515	935	15,976	mwt:01:sat:08
2	-0.2311	920	14,960	kmwt:0:01:sat:08
3	-0.2311	920	14,960	ymwt:2016:01:sat:08
4	-0.2311	920	14,960	kymwt:0:2016:01:sat:08
5	-0.2156	2,633	481,468	kym:0:2016:04
6	-0.2103	2,368,007	4,737,613	y:2016
7	-0.2099	9,277	110,093	kywt:0:2016:thu:08
8	-0.1983	0	2,889	kmwt:20:08:tue:13
9	-0.1967	0	2,159	kymwt:20:2017:08:tue:13
10	-0.1941	9,359	110,343	kwt:0:thu:08

に成功した。急ブレーキ識別に寄与した特徴量の組を調査したところ、急ブレーキの発生条件として、日時や曜日などの時間的要因が強く関係していることを明らかにした。

急ブレーキに関連する上位の特徴量を用いて、その特徴量に関連する行動の位置を地図上にプロットした。その結果、頻度ベースの手法で得られた位置とは異なることが判明した。つまり、急ブレーキは、車名、位置、走行方向などの車両関連要因だけでなく、年、月、曜日、時間帯などの時間的要因も関係していると考えられる。

表 5.8 急ブレーキの識別性能 (学習用データ・5 分割交差検証)

K	Precision	Recall	F 値	Accuracy
ALL	0.9166	0.9490	0.9325	0.9313
10	0.5325	0.9974	0.6941	0.5594
20	0.5596	0.9949	0.7163	0.6057
30	0.5560	0.9950	0.7131	0.5991
40	0.5657	0.9816	0.7176	0.6134
50	0.6419	0.9373	0.7620	0.7072
60	0.6953	0.8292	0.7171	0.7057
70	0.7635	0.7178	0.6968	0.7334
80	0.7380	0.8490	0.7895	0.7738
90	0.7808	0.7353	0.7217	0.7509
100	0.8527	0.5025	0.5722	0.6889
200	0.8889	0.9032	0.8960	0.8952
300	0.8915	0.9215	0.9063	0.9047
400	0.8948	0.9401	0.9169	0.9148
500	0.9014	0.9488	0.9245	0.9225
600	0.9086	0.9459	0.9269	0.9254
700	0.9076	0.9483	0.9275	0.9259
800	0.9066	0.9503	0.9279	0.9262
900	0.9101	0.9487	0.9290	0.9275
1000	0.9115	0.9481	0.9295	0.9281
5000	0.9154	0.9485	0.9317	0.9305
10,000	0.9161	0.9488	0.9321	0.9310

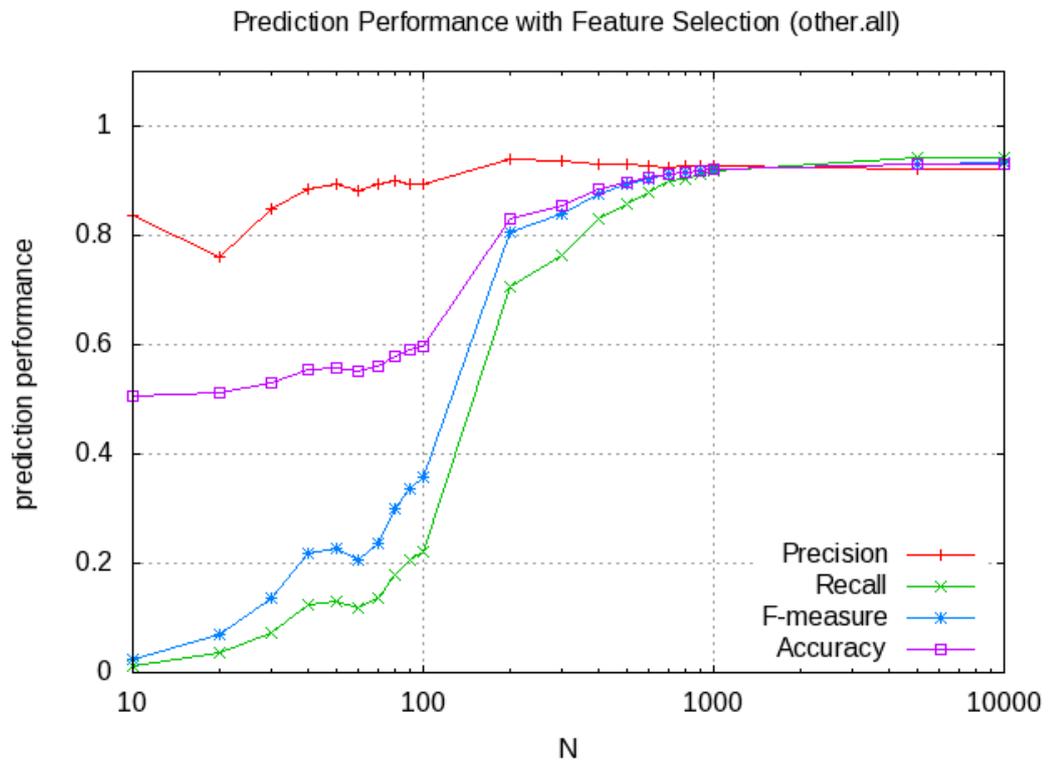


図 5.4 特徴選択を行った場合の急ブレーキ識別性能（地点の特徴量を除いたとき）

表 5.9 急ブレーキの識別性能（全データ・5 分割交差検証）

K	Precision	Recall	F 値	Accuracy
ALL	0.9212	0.9431	0.9320	0.9312
10	0.8353	0.0128	0.0248	0.5049
20	0.7606	0.0361	0.0690	0.5125
30	0.8481	0.0734	0.1351	0.5302
40	0.8857	0.1238	0.2171	0.5540
50	0.8931	0.1295	0.2259	0.5571
60	0.8813	0.1172	0.2068	0.5508
70	0.8939	0.1359	0.2358	0.5600
80	0.9000	0.1794	0.2986	0.5797
90	0.8943	0.2068	0.3355	0.5912
100	0.8929	0.2225	0.3562	0.5980
200	0.9406	0.7073	0.8074	0.8313
300	0.9372	0.7626	0.8409	0.8558
400	0.9308	0.8288	0.8769	0.8836
500	0.9304	0.8582	0.8929	0.8970
600	0.9280	0.8794	0.9030	0.9056
700	0.9255	0.8985	0.9118	0.9131
800	0.9284	0.9022	0.9151	0.9163
900	0.9272	0.9107	0.9189	0.9196
1,000	0.9268	0.9178	0.9223	0.9227
5,000	0.9211	0.9425	0.9317	0.9309
10,000	0.9211	0.9429	0.9319	0.9311

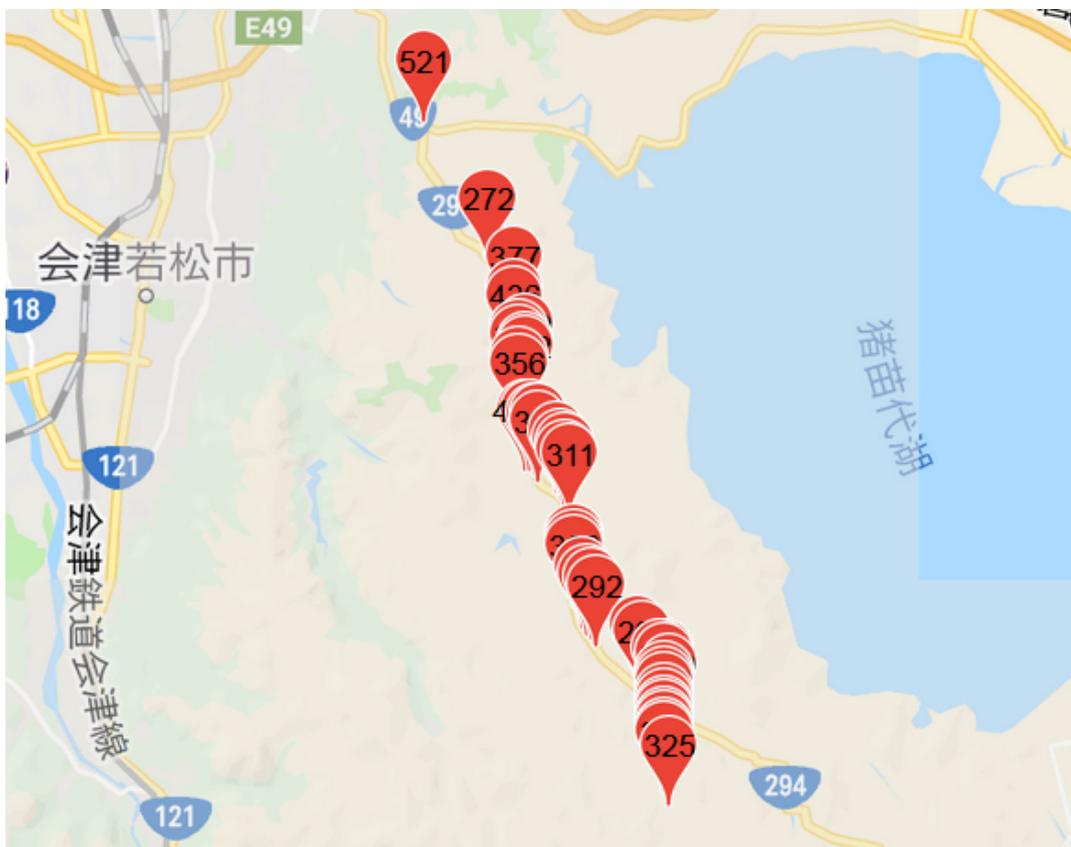


図 5.5 急ブレーキ正例の特徴語 “kym:10:2017:02” にあてはまる地点上位 50 件

第6章

まとめ・今後の課題

6.1 本研究の貢献

本研究は、人が移動するために利用する社会インフラのひとつとして不可欠な公共交通機関である路線バスに着目して、バス車両の位置情報履歴を大量に収集蓄積する手法や、それにより生成される位置情報ビッグデータの処理手法や活用手法を構築した。

まず、路線バス事業を取り巻く課題として「遅延の常態化」「利用客の減少」の2つに着目した。これらの課題は、様相や程度は違うものの、都市部と地方の両方において共通に見られるものであった。課題解決のためにはバスの運行状況を定量的に把握する必要がある、それを常時自動的に行うには情報システムを導入構築することが必須といえる。バスの運行状況を記録することに長けたシステムは従来より多数提案実装され、利用客への案内などとも連携した大規模なシステムが実用化されている。しかしながら、それらのシステムは一般的に高価であることから、特に地方の中小路線バス事業者において導入が困難であった。

そこで、本研究では次のアプローチにより課題解決を目指した。第一に、バス遅延の常態化問題を解決するために、遅延の問題を時間的傾向によって決まるような、運行ダイヤの調整によって解決可能な遅延と、日時によって変わり、運行ダイヤに反映しきれないか反映すると不便を招く遅延とに分けた。そして、そのうち後者を解決対象として、バスの運行所要時間を予測する手法を開発した。予測手法にはすでに多くの先行研究が存在するものの、それらのほとんどがリアルタイムな運行状況データの存在を要請しており、リアルタイムな情報収集はシステムのコストを高くする一因となる。その解決策として、データ収集コストを低くした場合に取得が困難となるリアルタイムな情報が存在しない場合において、リアルタイム性の低い情報を利用しつつ、リアルタイムな情報を利用した場合と遜色のない運行所要時間予測精度を達成することを目指した。その結果、予測に必要な情報と、機械学習アルゴリズムを選定し、その性能がリアルタイムな情報を利用する先行研

究と比べて遜色がないことを示した。第二に、リアルタイム性の低い情報を確実に収集できるような収集装置の設置手法・設置地点を定め、さらに、予測精度をより向上させることを目指す場合に収集用装置を増設する箇所の選定手法を示した。第三に、路線バス等の自動車の安全運行実現につながる要素のひとつとして急ブレーキに着目し、その発生要因を位置情報履歴データと機械学習を組み合わせる手法を提示した。

以下、各章の貢献を述べる。第2章では、位置情報履歴データ収集システムについて、本研究で採用している収集方式を他の収集方式との差異とともに示した。また、位置情報履歴データの処理手順を定式化し、位置情報履歴データのほか、バス停の設置位置情報のみがあれば、データ処理による運行状況算出、および運行所要時間予測のための特徴量算出が可能であると示した。

第3章では、まず、隣接するバス停間の運行所要時間を予測する方法を明らかにした。位置情報ビッグデータを用いてモデルを構築した結果、予測性能の向上には、予測対象区間の前の区間の所要時間、および予測対象区間に達するまでのバスの総停車回数を追加することで予測性能が向上することを確認した。しかしながら、この特徴量は常に得られるとは限らず、また、リアルタイムな情報であることから取得に大きなコストを要する。そこで、リアルタイム性が低くとも有用な特徴量として「予測対象区間の1時間前における、運行所要時間の平均値」を算出し、所要時間の予測を行った。その結果、リアルタイム性の低い情報を用いても、先行研究と遜色のない運行所要時間予測精度を達成した。

第4章では、位置情報履歴データを収集する設備について、低コストかつ高い予測精度を実現するための収集用路側機設置箇所の選定手法を示した。第3章で見出した特徴量「1時間前の運行所要時間の平均値」を実際に取得可能であるかを調べ、バス運行実績における運行時間帯数のうち、約75%の時間帯についてその特徴量が取得可能であるとの結果を得た。また、本研究における位置情報履歴収集用の路側機は、九州大学伊都キャンパス構内入口1箇所にも設置されていることから、この1箇所のみで1時間前の運行所要時間を算出するための情報が収集可能であるかを解析した。その結果、1箇所だけでは回収不可能な時間帯が存在したことから、1時間以内にバスが確実に通過する地点として、路線の端点である九大学研都市駅付近に路側機を追加設置することで収集できるようになると結論付けた。一方、全体の約25%の時間帯において1時間前の運行所要時間が得られなかったことから、その場合に予測精度の低下を避けるためには路線途中で路側機を増設すると効果があることを、予測対象時間帯における各バス停間の所要時間を特徴量として順次加えて予測精度を比較することにより示した。また、その際の選定基準として、特徴量重要度が3位以内となるバス停間の所要時間、その中でも起点側に近いものを得られるようなバス停間に設置することが望ましいとした。

第5章では、交通安全の要素として交通事故防止に着目し、事故の発生、または事故を回避した場面に現れる自動車の危険な走行挙動のひとつである「急ブレーキ」の発生条件

を抽出した。会津若松市公用車・公共交通車両走行情報オープンデータセットに対して、サポートベクターマシンを利用した機械学習と、特徴選択を組み合わせた手法により、急ブレーキ識別に寄与した特徴量の組を調査したところ、急ブレーキの発生条件として、日時や曜日などの時間的要因が強く関係していることを明らかにした。急ブレーキに関連する上位の特徴量を用いて、その特徴量に関連する挙動の位置を地図上にプロットした結果、急ブレーキ事象のデータ件数をもとにして得られた位置とは異なることが判明した。

本研究の手法、およびシステムは、特別な装置を利用せず、広く利用可能な位置情報収集・蓄積システムのみを利用して構築した。その際のコストは低廉であり、日本国内のみならず世界で広く適用可能である。また、本研究を通して、路線バスの運行改善に資する要素や知見を探ることができた。運行所要時間予測に重要な特徴量として時間的要因である「運行時間帯のうち、時の値」と「1時間前の運行所要時間の平均値」を見出した。また、運行所要時間予測に必要な位置情報ビッグデータが30分や1時間のように間欠的な収集しかできない場合でも、そのリアルタイム性の低い情報を利用しつつ先行研究と比してよい精度の運行所要時間予測手法を開発できた。さらに、予測のためのデータ整備に必須のものとして、位置情報ビッグデータの他は、バス停の位置情報を別途準備して照合すればよいようにシステムを構築できた。

6.2 本研究における限界

本研究で達成できていないこと、未解明のことは数多いが、その中でも代表的なものを2つ示す。

第一に、運行所要時間予測に有用で、かつリアルタイムな情報が得られない場合に利用できる情報として、リアルタイム性が下がるもののリアルタイムに近い情報である「予測対象区間における1時間前の運行所要時間実績値の平均値」を採用した。本情報の選定にあたり、対象路線である昭和バス九州大学線が時刻表上最大30分の所要時間であることから、その時点から余裕を持った時間間隔であるとして「1時間前」を基準として選択した。しかし、この値が本研究の目標である「高精度かつ低コストな運行所要時間予測」を達成する最適な値であるかの検証が十分でない。

第二に、本研究の対象路線と比べて、距離やネットワークの複雑度が増したバス路線へ拡張適用するにあたって、必要な追加条件の検討ができていない。データ収集用路側機の設置手法や機械学習モデルの構築手法、予測に用いる特徴量の妥当性を追究する必要がある。

6.3 今後の課題

本研究における路線バスの運行改善は、利用客の利便性や満足度を向上させることを主眼に置き、運行所要時間を予測する手法の研究開発を実施した。本研究の発展として、ネットワークの複雑度が増したバス路線や、その他の複数の交通機関に対する解析および所要時間予測への適用が考えられる。

また、バスの運行改善には、配車の効率化、勤務体系の最適化といった、運行事業者の業務改善の要素も重要である。地域公共交通を長期的に維持運営するには、利用客にとって使いやすいサービスを目指すのはもちろんのこと、運行事業者が事業運営を長く続けられるようにしたり、運行費用を低く抑えられるようにして財政的課題を解決する必要がある。今後、開発手法の実地検証の拡大や、システムとしての確立、バス事業の改善に重要な他の要素を活用するためのシステムやアルゴリズムの探求を今後の課題としたい。

謝辞

本論文の執筆にあたり、研究生から博士後期課程の四年間にわたり多大なる御指導御鞭撻ならびに叱咤激励を頂きました九州大学大学院システム情報科学研究所の峯恒憲准教授に深く御礼申し上げます。

本論文をまとめる過程において論文を御精読いただき、多くの御助言御指導を頂いた九州大学大学院システム情報科学研究所の荒川豊教授、ならびに川崎洋教授に深く感謝申し上げます。

本研究に際し、九州大学情報基盤研究開発センターの廣川佐千男名誉教授、伊東栄典准教授、九州工業大学大学院工学研究所の池永全志教授、鹿児島大学情報基盤統括センターの下園幸一准教授、および株式会社 JPIX 顧問の石田慶樹氏より多大なる御助言を頂き、研究を推進することができました。深く感謝いたします。

九州大学大学院システム情報科学研究所知的情報環境研究室、およびヒューマンフィリックシステム研究室の皆様には、本研究への有益なご意見を多数頂きました。また、同研究室秘書の福田優子氏、平仲友紀子氏、濱田千佳氏、吉元貴代氏、テクニカルスタッフの工藤百合子氏には研究室での活動にあたり、事務手続き等で大変お世話になりました。皆様に心より御礼申し上げます。

前職の会社を退職して博士後期課程へ進学するにあたり、諸事情を理解してくださった上で温かく送り出してくださった前職の上司や同僚、また、同期の友人に大変感謝いたします。生活面や学費の面などで多大なる支援を頂いた両親に感謝します。また、研究はもとより生活等さまざまな面において忌憚なき意見をくれた双子の姉に感謝します。

最後に、末の孫としてたくさん可愛がってくれ、博士後期課程への進学に際しても応援を頂き、私の研究生生活の途上で行年九十七歳の大往生を迎えた母方の祖母、深川アサヨに大いなる感謝を捧げ、墓前に成果を報告したいと思います。

参考文献

- [1] ITS japan - ITS とは. <https://www.its-jp.org/about/>. (Accessed on 2022.12.20).
- [2] 岡村健志, 永原三博, 熊谷靖彦, 菊池豊, 中川敏正. 地方部におけるバス案内システム「chi-bus」の提案と実践 岡村 健志. 自動制御連合講演会講演論文集 第 53 回自動制御連合講演会, pp. 315–315. 自動制御連合講演会, 2010.
- [3] 公共交通利用促進ネットワーク. <https://www.rosenzu.com/>. (Accessed on 2022.12.21).
- [4] 公共交通利用促進ネットワーク - 主なバスロケーションシステム製品リスト. <https://www.rosenzu.com/net/mieru/busloclist.html>. (Accessed on 2022.12.21).
- [5] 玉田匡, 伊藤昌毅, 川村尚生, 菅原一孔. バスの遅れ予測を目的とする, バスロケーションシステムから得た走行データの分析. In *IEICE Conferences Archives*. The Institute of Electronics, Information and Communication Engineers, 2013.
- [6] Mansur As and Tsunenori Mine. Prediction of travel time over unstable intervals between adjacent bus stops using historical travel time in both the previous and current time periods. In *Intelligent Transport Systems for Everyone's Mobility*, pp. 131–153. Springer, 2019.
- [7] Takuya Kawatani, Tsubasa Yamaguchi, Yuta Sato, Ryotaro Maita, and Tsunenori Mine. Prediction of bus travel time over intervals between pairs of adjacent bus stops using city bus probe data. *International Journal of Intelligent Transportation Systems Research*, Vol. 19, No. 2, pp. 456–467, 2021.
- [8] 内村圭一, 成松裕介, 衛藤旭秀, 胡振程. バスロケーション情報を用いたバス停間所要時間予測. IATSS review= 国際交通安全学会誌, Vol. 32, No. 3, pp. 224–231, 2007.
- [9] 今井瞳, 廣井慧, 河口信夫ほか. バス運行実績データの分析に基づく到着時刻予測モデルの提案と精度検証. 情報処理学会論文誌, Vol. 60, No. 1, pp. 101–117, 2019.
- [10] Cong Bai, Zhong-Ren Peng, Qing-Chang Lu, and Jian Sun. Dynamic bus travel time prediction models on road with multiple bus routes. *Computational intelli-*

- gence and neuroscience*, Vol. 2015, , 2015.
- [11] Johar Amita, SS Jain, and Pradeep Kumar Garg. Prediction of bus travel time using ann: a case study in delhi. *Transportation Research Procedia*, Vol. 17, pp. 263–272, 2016.
- [12] Mei Chen, Xiaobo Liu, Jingxin Xia, and Steven I Chien. A dynamic bus-arrival time prediction model based on apc data. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 19, No. 5, pp. 364–376, 2004.
- [13] Rafidah Md Noor, Ng Seong Yik, Raenu Kolandaisamy, Ismail Ahmedy, Mohammad Asif Hossain, Kok-Lim Alvin Yau, Wahidah Md Shah, and Tarak Nandy. Predict arrival time by using machine learning algorithm to promote utilization of urban smart bus, 2020.
- [14] 伊藤昌毅, 川村尚生, 菅原一孔. スマートフォンを利用したバスロケーションシステムの開発. 電子情報通信学会論文誌 D, Vol. 96, No. 10, pp. 2327–2339, 2013.
- [15] 公用車・公共交通車両走行情報 - データセット - data for citizen. <https://data.data4citizen.jp/dataset/10060158>. (Accessed on 2022.11.25).
- [16] BumpRecorder Web - 会津若松市オープンデータ (公用車・公共交通車両走行情報) . https://map.bumprecorder.com/open_data/aizuwakamatsu. (Accessed on 2022.11.25).
- [17] Creative Commons - 表示 4.0 国際 - CC BY 4.0. <https://creativecommons.org/licenses/by/4.0/deed.ja>. (Accessed on 2022.11.25).
- [18] 国土地理院「tky2jgd for windows ダウンロード」. https://www.gsi.go.jp/sokuchikijun/tky2jgd_download.html. (Accessed on 2022.11.30).
- [19] tky2jgd: 国土地理院提供測地系変換プログラム「tky2jgd」移植. <https://github.com/mugi-grainp/tky2jgd-shellscript>. (Accessed on 2022.11.30).
- [20] K Hubeny. Weiterentwicklung der gauss’ schen mittelbreitenformeln. *Z. Vermess*, Vol. 84, pp. 159–163, 1959.
- [21] アマノ技研ウェブサイト. <https://amano-tec.com/apps/paceruler.html>. (Accessed on 2022.11.30).
- [22] Trailnote : 2 地点間の距離の計算. https://www.trail-note.net/tech/calc_distance/. (Accessed on 2022.12.19).
- [23] 国土地理院「日本の測地系」. <https://www.gsi.go.jp/sokuchikijun/datum-main.html>. (Accessed on 2022.12.19).
- [24] 適切な誤差指標の選び方 - データサイエンス情報局. <https://analysis-navi.com/?p=2875>. (Accessed on 2022.11.29).
- [25] Sungil Kim and Heeyoung Kim. A new metric of absolute percentage error for

- intermittent demand forecasts. *International Journal of Forecasting*, Vol. 32, No. 3, pp. 669–679, 2016.
- [26] Pengfei Zhou, Yuanqing Zheng, and Mo Li. How long to wait? predicting bus arrival time with mobile phone based participatory sensing. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pp. 379–392, 2012.
- [27] Rui Zhang, Wenping Liu, Yufu Jia, Guoyin Jiang, Jing Xing, Hongbo Jiang, and Jiangchuan Liu. Wifi sensing-based real-time bus tracking and arrival time prediction in urban environments. *IEEE Sensors Journal*, Vol. 18, No. 11, pp. 4746–4760, 2018.
- [28] Zhao Huang, Qingquan Li, Fan Li, and Jizhe Xia. A novel bus-dispatching model based on passenger flow and arrival time prediction. *Ieee Access*, Vol. 7, pp. 106453–106465, 2019.
- [29] Simon J Berrebi, Etienne Hans, Nicolas Chiabaut, Jorge A Laval, Ludovic Leclercq, and Kari E Watkins. Comparing bus holding methods with and without real-time predictions. *Transportation Research Part C: Emerging Technologies*, Vol. 87, pp. 197–211, 2018.
- [30] Bin Yu, William HK Lam, and Mei Lam Tam. Bus arrival time prediction at bus stop with multiple routes. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 6, pp. 1157–1170, 2011.
- [31] Mei Chen, Xiaobo Liu, and Jingxin Xia. Dynamic prediction method with schedule recovery impact for bus arrival time. *Transportation Research Record*, Vol. 1923, No. 1, pp. 208–217, 2005.
- [32] Mathieu Sinn, Ji Won Yoon, Francesco Calabrese, and Eric Bouillet. Predicting arrival times of buses using real-time gps measurements. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pp. 1227–1232. IEEE, 2012.
- [33] Tsubasa Yamaguchi, AS Mansur, and Tsunenori Mine. Prediction of bus delay over intervals on various kinds of routes using bus probe data. In *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*, pp. 97–106. IEEE, 2018.
- [34] Tsubasa Yamaguchi, Ryotaro Maita, Takuya Kawatani, and Tsunenori Mine. Prediction of travel time over intervals between two bus stops using bus probe data. In *Proceedings of 17th ITS Symposium 2019*, 2019.
- [35] Jerome H Friedman. Greedy function approximation: a gradient boosting ma-

- chine. *Annals of statistics*, pp. 1189–1232, 2001.
- [36] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, Vol. 38, No. 4, pp. 367–378, 2002.
- [37] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.
- [38] Mansur As, Tsunenori Mine, and Hiroyuki Nakamura. Estimation of travel time variability using bus probe data. In *2017 6th IEEE International Conference on Advanced Logistics and Transport (ICALT)*, pp. 68–74. IEEE, 2017.
- [39] Mansur As and Tsunenori Mine. An adaptive approach for predicting bus travel time over unstable intervals. In *The 16th ITS Asia-Pacific Forum FUKUOKA, (ITS AP)*, pp. 146–160, 2018.
- [40] Mansur As and Tsunenori Mine. Dynamic bus travel time prediction using machine learning technique. *ACM IMCOM2018*, 2018.
- [41] Chi-Hua Chen. An arrival time prediction method for bus system. *IEEE Internet of Things Journal*, Vol. 5, No. 5, pp. 4231–4232, 2018.
- [42] Junbiao Pang, Jing Huang, Yong Du, Haitao Yu, Qingming Huang, and Baocai Yin. Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 9, pp. 3283–3293, 2018.
- [43] Ming Yang, Chao Chen, Lu Wang, Xinxin Yan, and Liping Zhou. Bus arrival time prediction using support vector machine with genetic algorithm. *Neural Network World*, Vol. 26, No. 3, p. 205, 2016.
- [44] 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 静岡県富士市の公共交通設計への最適化技術の活用. 都市計画論文集, Vol. 52, No. 3, pp. 1320–1326, 2017.
- [45] Yueying Huo, Wenquan Li, Jinhua Zhao, and Shoulin Zhu. Modelling bus delay at bus stop. *Transport*, Vol. 33, No. 1, pp. 12–21, 2018.
- [46] Frank Wilcoxon, SK Katti, and Roberta A Wilcox. *Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test*, Vol. 1. American Cyanamid Pearl River (NY), 1963.
- [47] Mansur As and Tsunenori Mine. Dynamic bus travel time prediction using an ann-based model. In *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication, IMCOM '18*, New York, NY, USA, 2018. Association for Computing Machinery.

-
- [48] Zegeye Kebede Gurmu and Wei David Fan. Artificial neural network travel time prediction model for buses using only gps data. *Journal of Public Transportation*, Vol. 17, No. 2, p. 3, 2014.
- [49] Hongjie Liu, Hongzhe Xu, Yu Yan, Zaishang Cai, Tianxu Sun, and Wen Li. Bus Arrival Time Prediction Based on LSTM and Spatial-Temporal Feature Vector. *IEEE Access*, Vol. 8, pp. 11917–11929, 2020.
- [50] Peilan He, Guiyuan Jiang, Siew-Kei Lam, and Dehua Tang. Travel-Time Prediction of Bus Journey With Multiple Bus Trips. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 11, pp. 4192–4205, November 2019.
- [51] Santa Maiti, Arpan Pal, Arindam Pal, Tanushyam Chattopadhyay, and Arijit Mukherjee. Historical data based real time prediction of vehicle arrival time. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1837–1842. IEEE, 2014.
- [52] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, Vol. 31, , 2018.
- [53] 古屋秀樹, 牧村和彦, 川崎茂信, 赤羽弘和. 車載型センサーを用いた車両挙動特異地点抽出に関する基礎的研究. *土木計画学研究・論文集*, Vol. 20, pp. 787–795, 2003.
- [54] 西堀泰英, 稲垣具志, 加知範康, 安藤良輔, 三村泰広. 自動車走行時の加速度発生状況と交通事故発生箇所の関連分析. *土木計画学研究・講演集*, Vol. 42, pp. 16–20, 2010.
- [55] Thorsten Joachims. *Learning to classify text using support vector machines*, Vol. 668. Springer Science & Business Media, 2002.
- [56] Hirotoshi Taira and Masahiko Haruno. Feature selection in svm text categorization. In *AAAI/IAAI*, pp. 480–486, 1999.
- [57] Thorsten Joachims. A support vector method for multivariate performance measures. In *Proceedings of the 22nd international conference on Machine learning*, pp. 377–384, 2005.

Appendix: 発表リスト

学術雑誌

1. Takuya Kawatani, Tsubasa Yamaguchi, Yuta Sato, Ryotaro Maita, Tsunenori Mine. *Prediction of Bus Travel Time over Intervals between Pairs of Adjacent Bus Stops Using City Bus Probe Data*. International Journal of Intelligent Transportation Systems Research volume 19, pages 456 – 467 (2021)
<https://doi.org/10.1007/s13177-021-00251-8>

学術会議発表

1. Takuya Kawatani, Eisuke Itoh, Sachio Hirokawa, Tsunenori Mine. *Machine learning and visualization of sudden braking using probe data*. 8th International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 67-72.
2. Takuya Kawatani, Eisuke Itoh, Sachio Hirokawa, Tsunenori Mine. *Location does not always determine sudden braking*. 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 875-882.
3. Takuya Kawatani, Tsunenori Mine. *Determination of collection points of bus probe data to achieve high prediction performance and low collection cost*. 2021 Thirteenth International Conference on Mobile Computing and Ubiquitous Network (ICMU), pp.1-6.

