

ランダムフォレストを用いた シェアサイクルの利用目的別トリップの推定

大井 啓史¹・野崎 脩人²・坂本 信³・浅田 拓海⁴・有村 幹治⁵

¹学生会員 室蘭工業大学 大学院工学研究科環境創生工学系専攻 (〒050-8585 室蘭市水元町27-1)

E-mail:21041011@mmm.muroran-it.ac.jp

²正会員 中央コンサルタンツ株式会社 (〒060-0034 札幌市中央区北4条東1丁目2-3)

E-mail: plzbo345@gmail.com

³正会員 株式会社ドーコン 防災保全部 (〒060-0001 札幌市中央区北1条西1丁目6)

E-mail:ss1948@docon.jp

⁴正会員 室蘭工業大学 大学院工学研究科 (〒050-8585 室蘭市水元町27-1)

E-mail:asada @mmm.muroran-it.ac.jp

⁵正会員 室蘭工業大学 大学院工学研究科 (〒050-8585 室蘭市水元町27-1)

E-mail: arimura@mmm.muroran-it.ac.jp

我が国では、自転車活用推進計画が閣議決定されるなど自転車の利用が推進される中でシェアサイクルの導入・利用が拡大しているが、その全利用者の利用目的を把握することは困難である。本研究は札幌市都市部のシェアサイクル・ポロクルのアンケートデータと利用実績データ、ポート周辺の建物施設データからなるデータセットにランダムフォレストを適用して利用目的を推定するモデルを構築し、得られたモデルを用いて全利用実績データの利用目的別トリップの判別を試みた。その結果、全利用実績データに対する利用目的別トリップの判別は、特に通勤・ショッピング目的に対して有用であることが示唆された。

Key Words : bike sharing, Random Forest, question survey, all usage data, trip purposes

1. はじめに

現在、わが国では、地球温暖化対策やコンパクトシティの形成、国民の健康増進等の観点から自転車の利用が推進されている。2018年6月には自転車活用推進計画¹⁾が閣議決定され自転車の活用に関する施策を総合的かつ計画的に推進することが図られた。その基本方針の一つとしてシェアサイクル施設の整備が重点的な検討項目として挙げられた。シェアサイクルは自転車を共同利用する交通システムであり、都市内に配置した各ポートにおいて、自転車の貸出・返却を無人管理で行う大規模なシステムの導入も進んでいる²⁾。シェアサイクルの導入意義としては、放置自転車の抑制、周遊性の向上による都市内活動の活性化、公共交通網との相互補完によるアクセシビリティの向上等³⁾が挙げられる。シェアサイクルを導入する都市は増加しており、2019年3月末には158都市で本格的に導入されている⁴⁾。

シェアサイクルの今後の発展や改善には利用者ニーズ

に合わせた運営が欠かせない。特に無人管理型シェアサイクルの場合、ポート間の移動記録データは常時観測・蓄積されるため、利用動向を把握するためにポート間ODや各種のデータマイニング等の適用が可能である。一方、主観的な情報である利用目的を自動的に収集することは困難であるため、シェアサイクルの利用パターンと利用目的を結びつけた分析は難しい課題であった。

本研究の分析対象とする札幌都心部で展開するシェアサイクル・ポロクル⁵⁾は札幌都心部の円滑な移動、にぎわい創出、CO2削減、路上駐輪の削減などへの寄与を目的に2011年から事業を開始している。2019年度からは運営時間延長や電動アシスト自転車の導入等により利用者が増加した。運営主体であるNPOポロクルは今後の需要増に備え利用者向けアンケートを2020年度に実施した。

そこで本研究では、上記の利用者アンケートデータと利用実績データを用いて、通勤や買い物等の利用目的別のトリップを推定するモデルを機械学習の一つであるランダムフォレストにより構築した。そのうえで構築した

モデルを2020年度の全利用実績データに適用し、従来把握することが困難であったシェアサイクルの利用目的別のトリップの推定を試みた。

本研究の構成は2章で既往研究のレビューを行い、3章でポロクルの概要説明と基礎集計を行う。4章でシェアサイクルの利用目的を推定するモデルの構築を行い、5章で得られたモデルを用いた利用目的の推定結果を示し、6章でまとめとする。

2. 既往研究

シェアサイクルの利用実績データに着目した定量的な研究としては、丸山ら³⁾はポロクルの全利用記録を用いて、可視化やパターン抽出・グラフマイニングによるアプローチで利用動態の把握を行っている。佐藤ら⁴⁾はコミュニティサイクルシステムの社会実験の利用実績データを用いて、利用状況を把握したうえで貸出頻度モデルを構築し、適切なステーション配置に関する定量的な分析をしている。南部ら⁷⁾は全国5事業のシェアサイクルの利用データを用いた変化動向の比較により公共交通の代替手段としての利用や休日の余暇利用が増加等のシェアサイクルにおけるCOVID-19の影響を確認している。

利用者に対するアンケートを用いた研究としては、橋本ら⁸⁾は岡山市のコミュニティサイクル「ももちやり」の利用者アンケート結果の分析からまちの魅力の要因として回遊しやすさに着目し、コミュニティサイクル導入がまちの魅力に与えることが把握している。

本研究の特徴は2020年度に実施された利用者アンケートデータとアンケート回答者の利用実績データを個人情報情報を除去した後に結び付けたうえで、都市計画基礎調査から利用ポート周辺の建物用途別延床面積を付与し、そのデータセットにランダムフォレスト法を適用し、利用者の利用パターンからその利用目的を推定するモデルを構築すること、また構築したモデルを2020年度の全利用実績データに適用し、全般的な利用目的別のトリップ数を把握することにある。

3. 概要と基礎集計

3.1 ポロクルの概要

札幌市のシェアサイクルであるポロクルの2020年度における概要を示す(表-1)。ポートは札幌市内の都市部を中心に48ヶ所に設置されている(図-1)。2011年度から導入されているポロクルだが、2019年度からNTTドコモと共同運営の施行が実施されており、新たに電動アシスト付き自転車の採用やNTTドコモが開発するアプリやAIによる再配置の最適化等のシェアサイクルシステムが導入された。営業時間の延長や保険補償の充実、貸出時の

表-1 2020年度ポロクル概要

シェアサイクル『ポロクル』概要	
自転車台数	400台
自転車車両	電動アシスト自転車
ポート数	48ヶ所
運営時間	24時間
利用体系	1回会員
	月額会員
	1日パス
	法人月額会員
会員登録方法・料金	アプリ・無料
貸出方法	カードキー
	パスコード
運営体制	認定NPO法人ポロクル NPO法人 ezorock ドコモ・バイクシェア
サポート	各種損害保険の付保

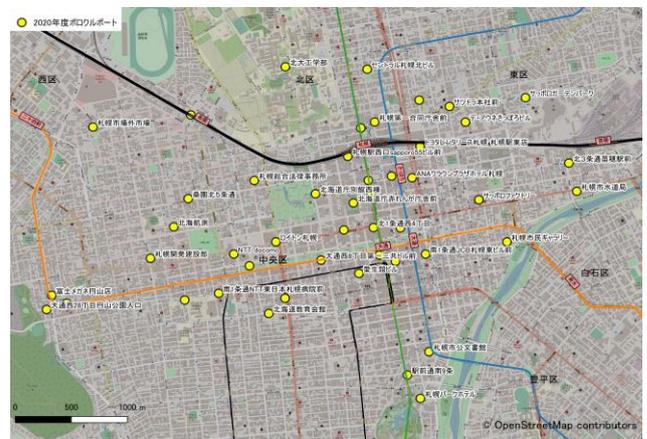


図-1 ポートマップ

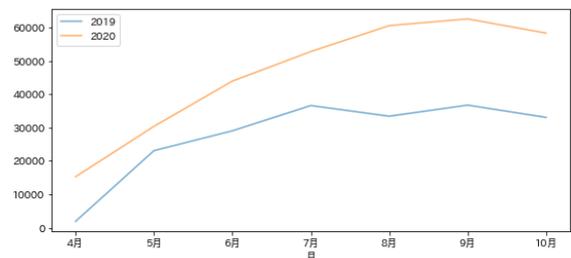


図-2 2019, 2020年度月別利用台数の推移

予約機能、コンビニエンスストアでの1日パス販売なども同年度に変更されている⁹⁾。

3.2 基礎集計

2019年度、2020年度のポロクル利用実績データから得られた利用台数に関する動向を示す。ポロクルの利用台数は2019年の利用は延べ193,896台、2020年度の利用は延べ323,766台であった。月別の利用台数は、2020年度のどの月においても2019年より利用台数が多く、2020年9月では62,550台で利用が最多となり、前年度の同月から約25,000台利用多くなった。月別の利用台数の推移では年度の営業開始する4月から夏にかけて増加し、営業が終了する10月末まで多く利用されている。(図-2)。

ポロクル2020年度アンケートに回答した利用者の利用目的について示す。ショッピング、通勤がほかの利用目的よりも多く、それぞれで全体の4分の1以上を占めている(図-3)。また、通勤目的とショッピング目的における利用開始時間の分布の特徴を示す(図-4、図-5)。通勤目的のピークは8時と17時前後にあり、通勤目的ではないピークはどちらも12時前後にある。ショッピング目的の突出したピークがなく昼間の利用回数が多い特有の分布となった。

4. 使用データ

4.1 使用データの概要

a) ポロクル2020年度全利用データ

ポロクルの2020年度全利用データは利用者数19,090人による323,766件の利用記録からなり多数の情報が含まれる。具体的には、ユーザーID、ユーザー種別、利用プラン、請求項目、請求額、利用開始日時、返却日時、利用時間(秒)、利用開始ポート名、返却ポート名、返却種別、車両情報が含まれている。なお冬季は積雪により営業が困難であるため、4月～10月までのデータである。本研究ではその中のユーザーID、請求項目、利用開始日時、返却日時、利用時間(秒)、利用開始ポート名、返却ポート名のデータを使用する。

b) ポロクル2020年度アンケートデータ

ポロクル2020年度アンケートデータは、ポロクル公式ホームページにて2020年度にポロクルを利用した人を対象に行ったアンケートデータであり、アンケート件数は1,194件である。アンケートデータには、年代・性別・住まいなどの個人属性・利用目的・公共交通機関との接続に関してなど計40個の回答データが含まれている。本研究では、まず「ポロクルの利用目的は?(2つまで選択可能)」という質問の回答で無回答と2つ回答しているデータを除いた386件のデータ抽出した。そしてその中でも通勤・通学・ショッピング・食事の回答データを対象にして利用目的を推定する。抽出したデータの中で通勤目的であると回答した利用者は121人、通学目的であると回答した利用者は3人、ショッピング目的であると回答した利用者は96人、食事目的であると回答した利用者は17人であった。

c) 都市計画基礎調査データ

都市計画基礎調査データは市街化区域内における全建物(430,884棟)について位置情報、建物用途、階数、構造および延床面積などが含まれている。本研究では2019年度の札幌市における「都市計画基礎調査」を使用した。その中でポートから半径100m以内にある合計15種類に区分される建物用途(表-2)と延床面積を用いる。

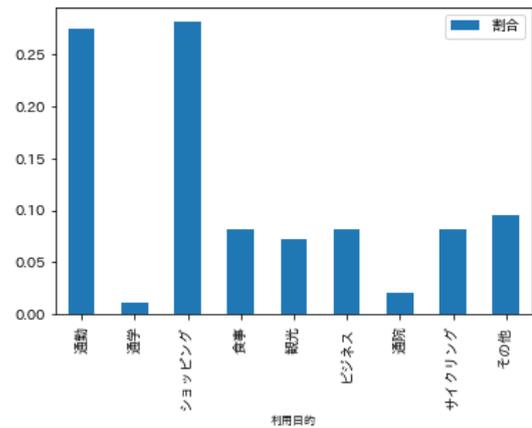


図-3 2020年度ポロクルの利用目的

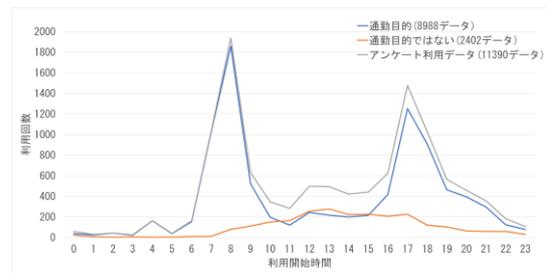


図-4 利用開始時間の分布

利用目的が通勤の場合

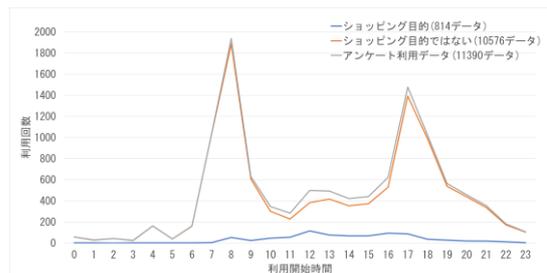


図-5 利用開始時間の分布

利用目的がショッピングの場合

表-2 建物分類

小分類	細分類
商業施設	-
官公庁施設	裁判所、道庁、税務署など
業務施設	会社、銀行、事務所など
宿泊施設	ホテル、旅館、モーテルなど
風俗娯楽施設	料理店、サウナ、ナイトクラブなど
遊技施設	カラオケ、パチンコなど
住宅	個人住宅
店舗併用住宅	-
共同住宅	公営
作業所併用住宅	-
文教厚生施設	学校、研究所、神社など
軽工業施設	-
サービス工業施設	-
運輸・倉庫施設	-
その他の施設	-

表-3 目的変数と説明変数

変数	変数名	単位
目的変数1	利用目的ダミー(通勤)	無次元
目的変数2	利用目的ダミー(通学)	無次元
目的変数3	利用目的ダミー(ショッピング)	無次元
目的変数4	利用目的ダミー(食事)	無次元
説明変数	利用時間	秒
	平日祝日ダミー	無次元
	利用開始時時間	秒
	返却時間	秒
	ポート間距離	km
	利用開始ポート半径100m以内の建物用途別延べ床面積(表-1全て)	m ²
	返却ポート半径100m以内の建物用途別延べ床面積(表-1全て)	m ²

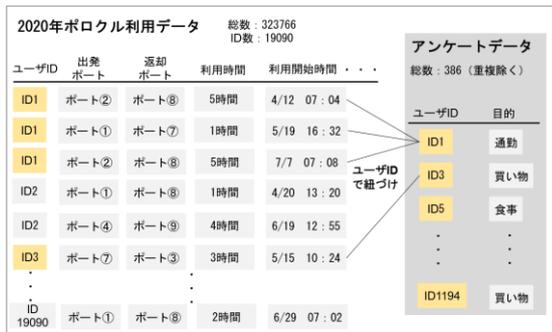


図-6 ユーザーIDによるデータの結びつけ

表-4 テストデータの精度

目的	テストデータの精度
通勤	92.2%
通学	99.8%
ショッピング	95.5%
食事	99.3%

表-5 アンケート利用データの利用目的と全利用データの推定した利用目的

目的	アンケート利用データ	推定データ
通勤	78.9%	64.7%
通学	0.5%	0.7%
ショッピング	7.1%	3.9%
食事	1.5%	0.5%

4.2 RF(ランダムフォレスト)による推定

本研究で用いたRFは複数の決定木から森を構成して、推定や識別することができる機械学習アルゴリズムである。特徴として、個々の決定木は高い識別性を持たないが、それらを複数用いることで高い予測性能を得られること、目的変数を推定する際の説明変数の寄与度を算出できること、説明変数の正規化の必要性がないこと等が挙げられる。推定精度は決定木の個数、決定木の最大高さ、パラメータの調整、ノードの分割関数、学習時の目的関数、特徴選択によって左右される。また、過学習などに注意する必要がある¹⁰⁾。

本研究では、まず4.1 a)で示した2020年度全利用データから、4.1 b)で示した2020年度アンケート回答者のユーザーIDにより抽出した11,390件のデータを用いる(以下、アンケート利用データと記す)。上記のユーザーIDによる結びつけのイメージを示す(図-6)。全ての利用記録について、利用時の平日・休日ダミー変数を設定した。また利用開始ポートと返却ポートの経度・緯度からポート間距離を算出して利用データに付与した。また利用開始ポートの半径100m以内の建物用途別延床面積と返却ポートの半径100m以内の建物用途別延床面積を各

利用開始ポート・返却ポートのデータに付与した。

上記で示したデータセットを学習モデルとして使用し利用目的の推定を行った。データセットを無作為に分割し、7割を学習データ、3割をテストデータとして使用する。学習と利用目的推定に使用する目的変数と説明変数を示す(表-3)。

4.3 精度検証と寄与度

テストデータの各利用目的の精度について検証した結果を示す(表-4)。精度は全ての利用目的で90%以上と高く、利用目的を十分に推定できる学習モデルを構築できたことがわかる。

推定した2020年度全利用データの各利用目的の寄与度を示す(図-7, 図-8, 図-9, 図-10)。この結果より、利用目的が通勤・ショッピング・食事の場合の利用目的を推定するのに寄与する説明変数は、利用時間(秒)・ポート

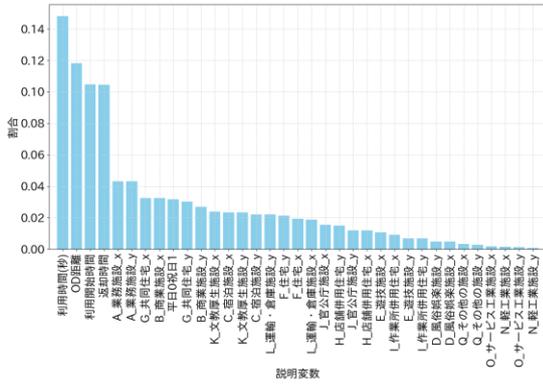


図-7 通勤目的の寄与度

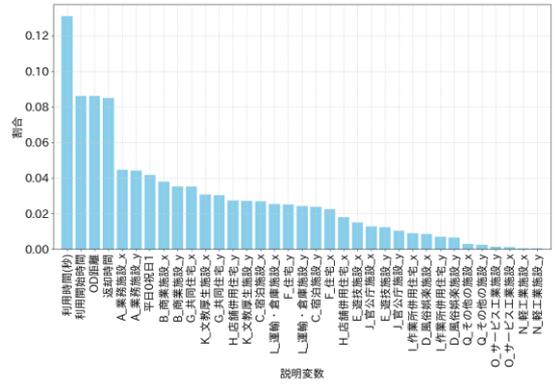


図-9 ショッピング目的の寄与度

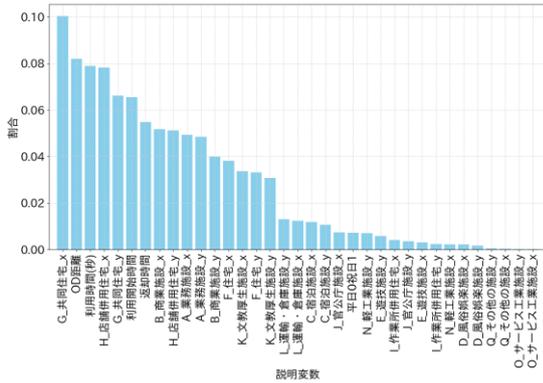


図-8 通学目的の寄与度

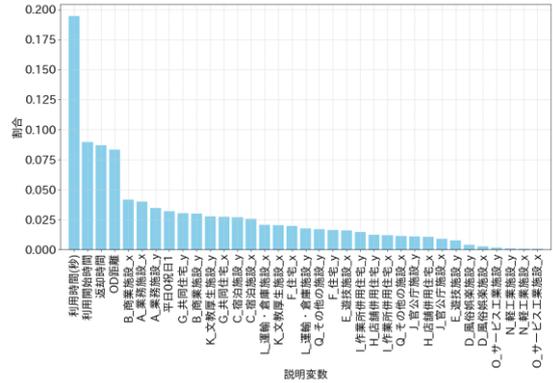


図-10 食事目的の寄与度

間距離・利用開始時間・返却時間であることが分かった。また、利用目的が通学の場合、利用開始ポートの半径 100m 以内の共同住宅の延床面積・ポート間距離・利用時間(秒)・返却ポートの半径 100m 以内の共同住宅の延床面積が寄与している。

5. 全利用実績データへの適用

アンケートデータから構築したRFを用いて2020年度ポロクル利用実績データの各利用目的別のトリップ数の推定を行う。アンケートデータと2020年度ポロクル利用実績データの推定結果を比較した(図-11)。その他の割合については、他4つの目的以外で示している。

2020年度ポロクル利用実績データにおける通勤目的の総トリップ数は209,477トリップとなった。通学は2,266トリップ、ショッピングは12,627トリップ、食事目的1,619トリップとなった。

また、2020年度全利用データの推定した各利用目的の利用時間を示す(図-12, 図-13)。3.2に示したアンケート利用結果との比較を以下に示す。利用目的が通勤である図-4と図-12を比較すると波形が似ており、通勤目的のピークはどちらも8時と17時前後にあるため、特徴を押さえて通勤目的を推定できているといえる。同様に、利用目的がショッピングである図-5と図-13を比較する

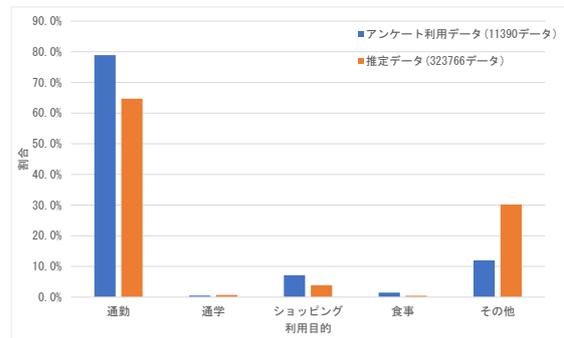


図-11 アンケート利用データの利用目的と全利用データの推定した利用目的の比較

と、どちらも突出したピークは見られず、ショッピング特有の昼間の利用回数が高い分布となった。

以上から、通勤、ショッピング目的のトリップを概ね正確に推定できたといえる。通学・食事目的では、データ数が少なかったため良い推定結果が得られなかったが、今後その利用目的での利用者に絞ったアンケートの実施等でデータを得ることにより推定モデルの精度向上が期待できる。

6. まとめ

本研究では、シェアサイクルのアンケートデータと利用実績データを結び付けたデータセットにRFを用いてシェアサイクル利用目的を推定するモデルを構築し、そのモデルを全利用実績データに適用した利用目的別トリップ数の推定を行った。その結果、1) 構築したモデルにより利用目的別の総トリップ数を得られた。2) アンケートで利用目的としての回答が多かった通勤・ショッピング目的では利用開始時間での特徴を押さえた推定が可能であった。3) データ数が少なかった通学・食事目的では、良い推定結果を得ることができなかった。

今後の課題として、利用目的を推定する際に説明変数として使用する建物用途別延床面積の範囲の拡大やRF以外の分類方法でのモデルの構築、交差検証の見直し等を行い、精度を高めることが挙げられる。

謝辞：本研究では、特定NPO法人ポロクルからの貴重なデータを頂いた。ここに記して感謝の意を表する。

参考文献

- 1) 国土交通省：自転車活用推進計画について、
<https://www.mlit.go.jp/common/001267507.pdf>
- 2) 一般社団法人日本シェアサイクル協会ホームページ、
<https://www.gia-jsca.net/sharecycle.shtml>, (2020/2/28閲覧)
- 3) データマイニングアプローチによるコミュニティサイクルの利用動態の抽出, 土木学会論文集D3(土木計画学), Vol.70, No.5, pp.671-680, 2014
- 4) 国土交通省：2019.11.6 全国シェアサイクル会議資料「シェアサイクルの取組等について」,
<https://www.mlit.go.jp/toshi/content/001384872.pdf>
- 5) ポロクルホームページ：<https://porocle.jp/>, (2020/12/15閲覧)
- 6) 佐藤仁美, 酒井良輔, 三輪富生, 森川高行：コミュニテ

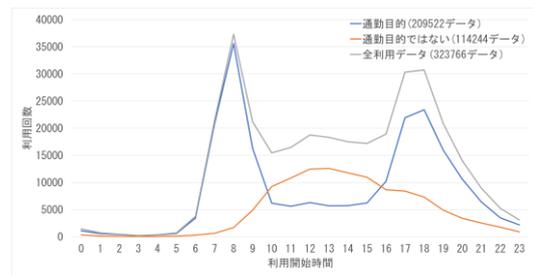


図-12 推定結果における利用開始時間の分布
利用目的が通勤の場合

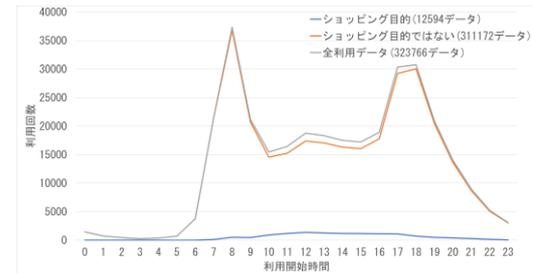


図-13 推定結果における利用開始時間の分布
利用目的がショッピングの場合

ィサイクルの利用実態とステーション配置に関する研究, 土木学会論文集D3, 69巻5号, L563-L570, 2013.

- 7) 南部浩之, 片岸将広, 熊谷美香子, 三浦清洋, 成嶋良太, 門脇照：我が国のシェアサイクル事業におけるCOVID-19の影響分析, 土木計画学研究発表会・講演, 2020
- 8) 橋本成仁, 中島那枝：コミュニティサイクルの導入がまちの魅力に与える効果に関する研究, 都市計画論文集 Vol.52 No.2, 2017
- 9) 認定 NPO 法人ポロクル 2019 年度事業レポート, https://porocle.jp/npo_porocle/pdf/porocle_report2019.pdf, (2020/2/28 閲覧)
- 10) 波部齊：ランダムフォレスト, 情報処理学会研究報告, Vol.2012-CVIM-182 NO.31, 2012

ESTIMATING TRIP PURPOSES OF A SHARED BICYCLE SYSTEM APPLYING A RANDOM FOREST

Hirofumi OI, Shuto NOZAKI, Shin SAKAMOTO,
Takumi ASADA and Mikiharu ARIMURA