

# 機械学習モデルを用いた シェアサイクルシステム利用目的の推定

丹度 彪雅<sup>1</sup>・大井 啓史<sup>2</sup>・浅田 拓海<sup>3</sup>・有村 幹治<sup>4</sup>

<sup>1</sup>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:23041052@mmm.muroran-it.ac.jp

<sup>2</sup>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

<sup>3</sup>正会員 室蘭工業大学助教 大学院工学研究科 もの創造系領域 (〒050-8585 室蘭市水元町 27-1)  
E-mail: asada@mmm.muroran-it.ac.jp

<sup>4</sup>正会員 室蘭工業大学教授 大学院工学研究科 もの創造系領域 (〒050-8585 室蘭市水元町27-1)  
E-mail: arimura@mmm.muroran-it.ac.jp.

近年、日本を含む多くの国でシェアリングサイクルのサービスが拡大している。シェアサイクルの今後の発展や、利用状況の改善には利用者ニーズの把握は必要不可欠である。一方で、無人管理シェアサイクルが多い日本においては利用者ニーズの収集は困難である。本研究では、利用実績データとアンケートデータを組み合わせ、ランダムフォレストモデル、XGBoost モデル、ロジスティック回帰モデルを適用することで、利用目的を推定するモデルを構築した。また、複数年度の利用実績データにモデルを適用することで、利用者の行動変化を評価した。その結果、距離とサイクリング時間が重要な説明変数であることがわかった。

**Key Words:** bike sharing, machine learning, question survey, all usage data, trip purposes estimate

## 1. はじめに

現在、我が国では、地球温暖化対策、災害時における交通機能の維持、国民の健康増進等様々な観点から自転車活用が促進されている。2017年5月には自転車の活用の推進に関する総合的かつ計画的な推進を図るため、我が国の自転車の活用の推進に関して基本となる計画として自転車活用推進法が施行された<sup>1)</sup>。2018年に策定された第1次自転車活用推進計画から、コロナ禍における生活様式・交通行動の変容、情報通信技術の発展等の社会情勢の変化を踏まえ、持続可能な社会の実現に向けた自転車の活用の推進を図るため2021年5月には第2次自転車活用推進計画が策定された。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挙げられている。シェアサイクルとは自転車を共同利用する交通システムであり、不特定多数の者の移動手段としての機能を有する、公共交通機関を補完する手段としての役割を担う、面的なネットワークの形成により自由度の高い移動が可能といった特性がある。シェアサイクルへの関心は高まりを見せており、2020年3月末時点で164都市が本格導入、54都

市が社会実験実施、40都市が導入検討中となっている。

シェアサイクルの国内における導入状況を見ると、約6割の都市がICTの活用による無人管理を行っており、自転車台数、ポート数で割合を見るとどちらも約8割が無人管理となっている<sup>2)</sup>。シェアサイクルの今後の発展や利用状況の改善には利用者ニーズの把握は不可欠である。しかし、無人管理型シェアサイクルの場合、移動記録データから利用動向の把握が可能である一方で、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。そのため、シェアサイクルの利用パターンと利用目的を結びつけての分析は困難であった。

本研究の対象とする、札幌市都心部でシェアサイクルとして提供されているポロクルは、民間が担う新たな公共的な交通サービスとしての事業展開、まちづくりへの貢献、自転車ルール・マナー、環境負荷低減に関する啓発等を目的として、2011年より事業を開始している。2019年度からは営業時間の拡大や、全自転車が電動アシスト機能付きに変わったことで利用者が増加した。また運営主体であるNPO法人ポロクルは、管理運営やサービス向上のためシーズンごとにアンケート調査を実施し

ている<sup>3)</sup>。

## 2. 先行研究のレビューと本研究の目的

シェアサイクルの利用データに関する先行研究として、南部ら<sup>4)</sup>は「ポロクル」、「まちのり」、「ももちやり」、「さいたま市コミュニティサイクル」、「Suicle」の5事業を対象とし、利用データから COVID-19 状況下における公共交通の代替手段としてのシェアサイクルの影響を分析している。

利用者アンケートを用いた先行研究として、橋本ら<sup>5)</sup>は岡山市コミュニティサイクル「ももちやり」を対象として、利用者アンケートから、利用者の「ももちやり」導入によるまちの変化に関する意識、利用前後での行動変化を分析し、コミュニティサイクル導入がまちの魅力に与える影響を示した。

大井ら<sup>6)</sup>の先行研究では札幌市内のシェアサイクル「ポロクル」を対象として、アンケートデータと利用実績データにランダムフォレストを適用することで、利用目的推定するモデルを構築している。しかし、利用目的によっては精度が十分でないこと、構築したモデルを用いて複数年度にかけての利用行動変化を把握できていない等の課題がある。

そこで本研究では、利用者アンケートデータと利用実績データを用いて、通勤とその他に分けた利用目的をトリップごとに推定するモデルをランダムフォレスト、XGBoost、ロジスティック回帰の3つの手法を用いて構築した。そのうえで構築したモデルを2019年度から2022年度までの利用実績データに適用し、利用者の複数年度にかけての行動変化の把握を試みた。また利用目的の推定に寄与する特徴量について機械学習モデルを解釈するSHAP指標を用いて分析する。

## 3. 概要と使用データ

### (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2022年度における概要を表-1に示す。ポートは札幌中心市街地に53カ所設置されている。2011年度から営業されているポロクルだが、2019年度からは第二期としてNTTドコモと共同運営が実施されており、普通自転車から電動アシスト付き自転車への転換、営業時間の延長等が行われた。

### (2) 使用データ

#### a) ポロクル2019年度-2022年度全利用データ

表-1 2022年度ポロクル概要

| 「ポロクル」概要 |                                 |
|----------|---------------------------------|
| 認証システム   | 車載端末認証                          |
| 自転車車両    | 電動アシスト自転車                       |
| 自転車台数    | 480台                            |
| ポート数     | 53カ所                            |
| 料金プラン    | 月額会員                            |
|          | 1回会員                            |
|          | 1日パス                            |
|          | 法人月額会員                          |
| 利用時間     | 24時間                            |
| 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                      |
| その他      | 損害保険の付保                         |
|          | 「ドコモ・バイクシェア-バイクシェアサービス」アプリが利用可能 |

2023年2月末現在

ポロクルの2020年度全利用データをモデル構築に使い、他年度の全利用データは利用行動変化を把握するために用いる。2020年度全利用データは323,766トリップ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、料金プラン、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、車両情報等が含まれる。なお冬季は積雪により営業が困難なため4月~11月までのデータである。本研究ではその中のユーザーID、料金プラン、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年度におけるポロクル利用者を対象に行なったアンケートデータであり、アンケート件数は1194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328件のデータを抽出した。その中で通勤と通勤以外の回答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117人、通勤目的以外であると回答した利用者は211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。

(3) データセットの概要

本研究では、(2)a)で示した 2020 年度全利用データと(2)b)で示した2020年度アンケートデータをユーザーIDで紐づけして算出した 9,995 件のデータ（以下、アンケート利用データと記す）を作成した。アンケート利用データを、料金プランの月額会員 7,605 件、1回会員 2,390 件に分けて利用する。アンケート利用データ月額会員の中で通勤目的であると回答した利用パターンは 6,950 件、通勤目的以外と回答した利用パターンは 655 件であった。アンケート利用データ 1回会員の中で通勤目的であると回答した利用パターン 1,230 件、通勤目的以外と回答した利用パターンは 1,160 件であった。アンケート利用データの利用目的別の分布を示す。（図-1）それぞれの利用目的と平日か休日であるかダミー変数を用いて分類し、全てのパターンで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する。これらのデータに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利用開始ポートと返却ポートで紐づけることで、データセットを構築した。

表-2 建物分類

| 小分類      | 細分類         |
|----------|-------------|
| 商業施設     | -           |
| 官公庁施設    | 裁判所,道庁など    |
| 業務施設     | 会社,事務所など    |
| 宿泊施設     | ホテル,旅館など    |
| 風俗娯楽施設   | 料理店,クラブなど   |
| 遊技施設     | カラオケ,パチンコなど |
| 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,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   | -           |

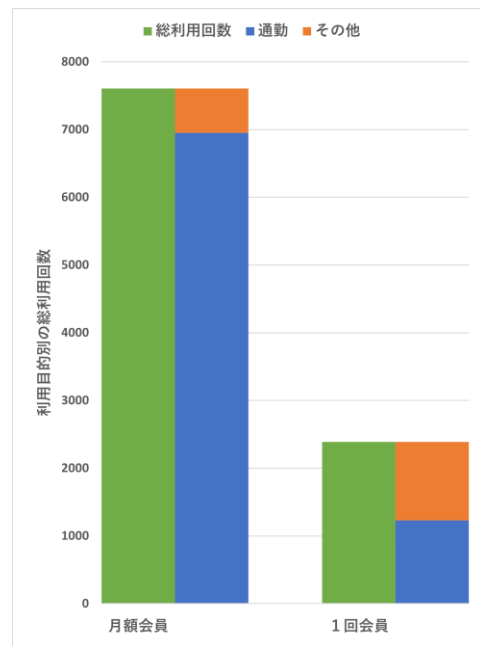


図-1 2020年度ポロクルの利用目的

4. 機械学習モデルによる推定と解釈

(1) 目的変数と説明変数

本研究では、目的変数、説明変数を表-3 のように設定することで、2020 年度のアンケートデータから、2020 年の全利用データにはない、各利用目的の割合を推定する。さらに、各説明変数の目的変数に対する影響度を表す寄与度を求めることにより、要因分析も行う。この分析に適した手法を検討するため、3つの手法を用い、そ

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(その他)                      | 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-2全て) | m <sup>2</sup> |
| 返却ポート半径100m以内の建物用途別延べ床面積(表-2全て) | m <sup>2</sup>                    |                |

の精度について比較を行う。

## (2) 機械学習モデル

### a) ランダムフォレスト

ランダムフォレストとは、バギング（全体からランダムに取り出した一部のデータから独立に多数のモデルを作成し集約する手法）をベースとして、決定木を用いる手法である。特徴として、個々の決定木機は高い精度を持たないが、複数用いることで高い予測精度を得られること、目的変数を推定する際の説明変数の重要度を出力可能であること、決定木を使用するため非線形な事象に対応できることがあげられる<sup>7)</sup>。ハイパーパラメータとして、決定木の個数、決定木の最大高さ、ノードの分割関数等がある。

### b) XGBoost

XGBoost とは、ブースティング（一部のデータを繰り返し抽出し、逐次的にモデルを学習させる手法）をベースとして決定木を用いる手法である。特徴として木の構造を複雑になりすぎないように正則化項をつけて形が固定された木構造の最適解を近似的に求めていく、木構造は木の分岐する前とした後の誤差を考慮して、構造を決めていくことなどがあげられる<sup>8)</sup>。ハイパーパラメータとして、決定木の深さ、決定木の葉の重みに関する正則化項、ランダムに抽出される標本の割合等があげられる。

### c) ロジスティック回帰

ロジスティック回帰とは線形回帰の出力にシグモイド関数を用いることで、説明変数から 2 値の目的変数がおこる確率を予測する手法である。特徴として説明変数間に強い相関がある場合には、回帰係数が安定せず信頼性が低くなる等の問題がある<sup>9)</sup>。

## (3) 機械学習モデルの解釈手法 (SHAP)

SHAP は、精度は高いが解釈性が低い深層学習等のモデルを解釈説明するための手法の一つである。機械学習で学習したモデルを単純なモデルで近似し、可読表現を用いて説明する<sup>8)</sup>。協力ゲーム理論の Shapley Value を機械学習に応用したものである。具体的には協力ゲームが報酬により限界貢献度が定義されるのに対し、SHAP では予測値を用いて貢献度を計算する。各特徴量がある時、ない時の予測値の差分をもって限界貢献度をとり、すべての順番に対し求めて平均を取ることで、説明変数の重要度を計算する。

## (4) 学習データとパラメータ調整

予測モデルの精度を検証するため 3.(3)で示し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用いる。また、通勤とその他でサンプル数に差があるため Python の SMOTE ライブラリを用い

てオーバーサンプリングを行った。これにより、サンプル数の少ないクラスに対するモデルの感度が低くなることを、対象データの不均衡を解消することで改善できる。用いた学習データの中で、月額会員は通勤：その他が 4,865 : 458 と約 10 倍の差があったが、オーバーサンプリングを行い 4,865 : 4,865 にすることで偏りをなくした。1 回会員に関してもオーバーサンプリングを行い、通勤：その他が 871 : 802 だったものを 871 : 871 と修正した。ランダムフォレスト、XGBoost ではハイパーパラメータの調整が精度に大きく影響するため、ベイズ最適化を用いてパラメータの調整を行った。

## 5. 結果

### (1) 精度検証

テストデータの各利用目的における推定精度検証結果を示す (表-4)。

テストデータの精度に関して、月額会員の場合はランダムフォレスト、XGBoost とともに 97%以上、1 回会員の場合は、ランダムフォレスト、XGBoost とともに 91%以上と高い数値を示した。ランダムフォレスト、XGBoost を用いることで、利用目的を十分に推定できる学習モデルを構築できた。

表-4 テストデータの精度検証

| 機械学習モデル   | テストデータの精度 |       |
|-----------|-----------|-------|
|           | 通勤        | その他   |
| 月額会員      |           |       |
| ランダムフォレスト | 97.2%     | 97.2% |
| XGBoost   | 97.1%     | 97.0% |
| ロジスティック回帰 | 80.7%     | 80.7% |
| 1回会員      |           |       |
| ランダムフォレスト | 91.5%     | 91.5% |
| XGBoost   | 91.5%     | 91.4% |
| ロジスティック回帰 | 69.9%     | 69.9% |

### (2) 2019年度-2022年度の利用目的推定

#### a) ランダムフォレストモデルでの推定

ランダムフォレストを用いて 2019 年度から 2022 年度のポロクル全利用データに対して、利用目的を推定した結果を示す。(図-2)

図-2 より月額会員において通勤目的の利用者数は 2019 年度から 2021 年度にかけて減少傾向にあるが、2022 年度には増加している。一方、その他の利用者数は 2019 年度から 2021 年度では増加傾向にあるが、2022 年度に

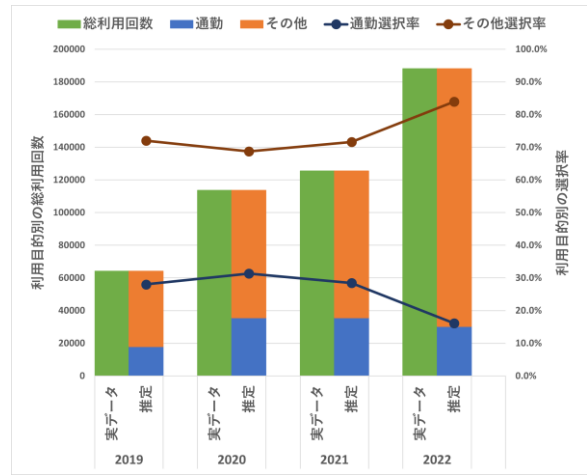
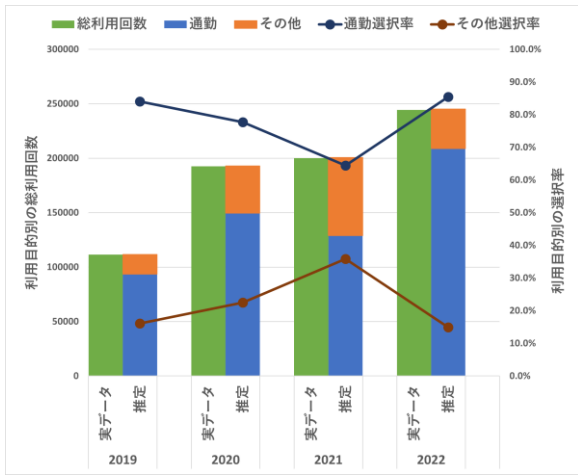


図-2 実際の利用データとランダムフォレスト推定データの比較 (左：月額会員 右：1回会員)

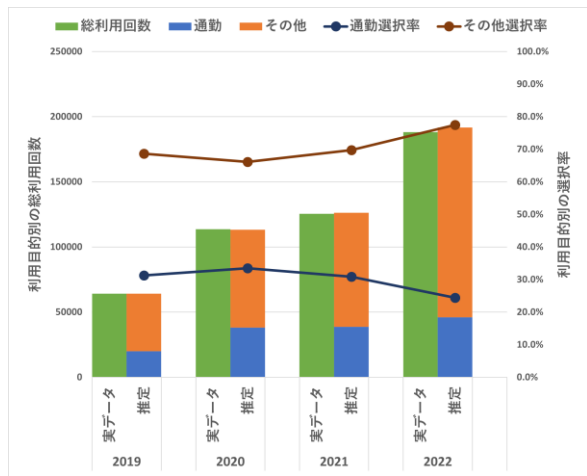
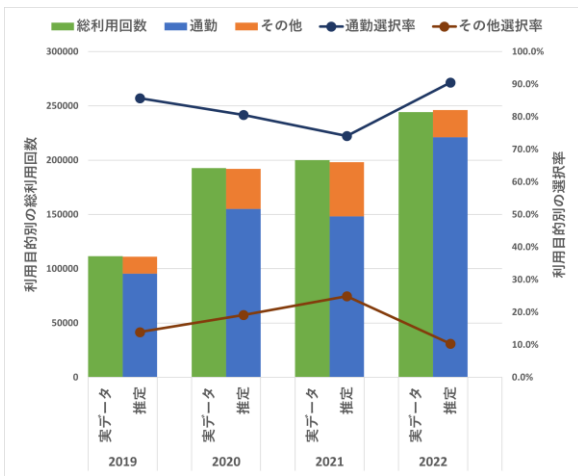


図-3 実際の利用データと XGBoost 推定データの比較 (左：月額会員 右：1回会員)

は減少していることが分かる。1回会員において通勤目的の利用者数は 2020 年度以降減少傾向にあり、その他の利用者数は 2020 年度以降増加傾向にある。

b) XGBoost モデルでの推定

XGBoost を用いて 2019 年度から 2022 年度のポロクル全利用データに対して、利用目的を推定した結果を示す (図-3)。

図-3 より XGBoost での推定結果は、ランダムフォレストでの推定結果と同様の傾向が見られた。

(3) 利用目的推定の解釈

2020 年度データを用いた月額会員の通勤目的利用結果に関して、機械学習モデルの解釈のため Shapley Value をランダムフォレスト、XGBoost とともに計算した。

説明変数の寄与度と目的変数の相関を示す (図-4, 図-5, 図-6)。図-4, 図-5 は横軸が目的変数の値で縦軸が説明変数の貢献度の高さ、色が各サンプルの特徴量の値 (赤が正、青が負) である。図-6 は OD 距離を横軸、

Shapley Value を縦軸に示した図である。利用目的が通勤の場合、月額会員、1回会員ともに、OD 距離、返却時間、利用開始時間、利用時間 (秒) が重要な説明変数であると分かる。

図-4 よりランダムフォレストによる利用目的の推定では、OD 距離を見ると青 (OD 距離が短いとき) が正の方向、赤 (OD 距離が長いとき) が負の方向へ多く分布している。OD 距離は距離が短いほどその他が選択される方向へ作用する負の相関がみられた。図-6 左図からも負の相関があることが分かる。

図-5 より XGBoost による利用目的の推定では、OD 距離を見ると赤 (OD 距離が長い) ほど正の方向、青 (OD 距離が短い) ほど負の方向へ多く分布しており、OD 距離が長いほど通勤目的が選択される方向へ作用する正の相関がある。図-6 右図からも同様のことがいえる。

これらより、ランダムフォレストと XGBoost は推定に重要な説明変数は同じであるものの、相関は逆を示す事を確認した。

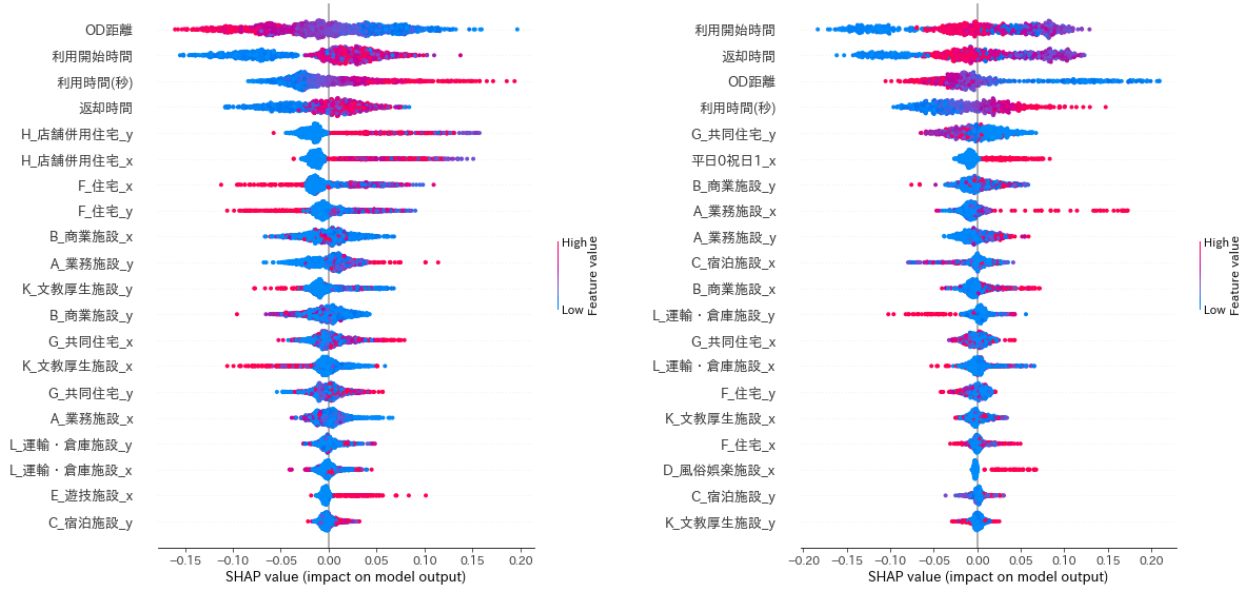


図-4 通勤目的の Shapley Value ランダムフォレスト (左：月額会員 右：1回会員)

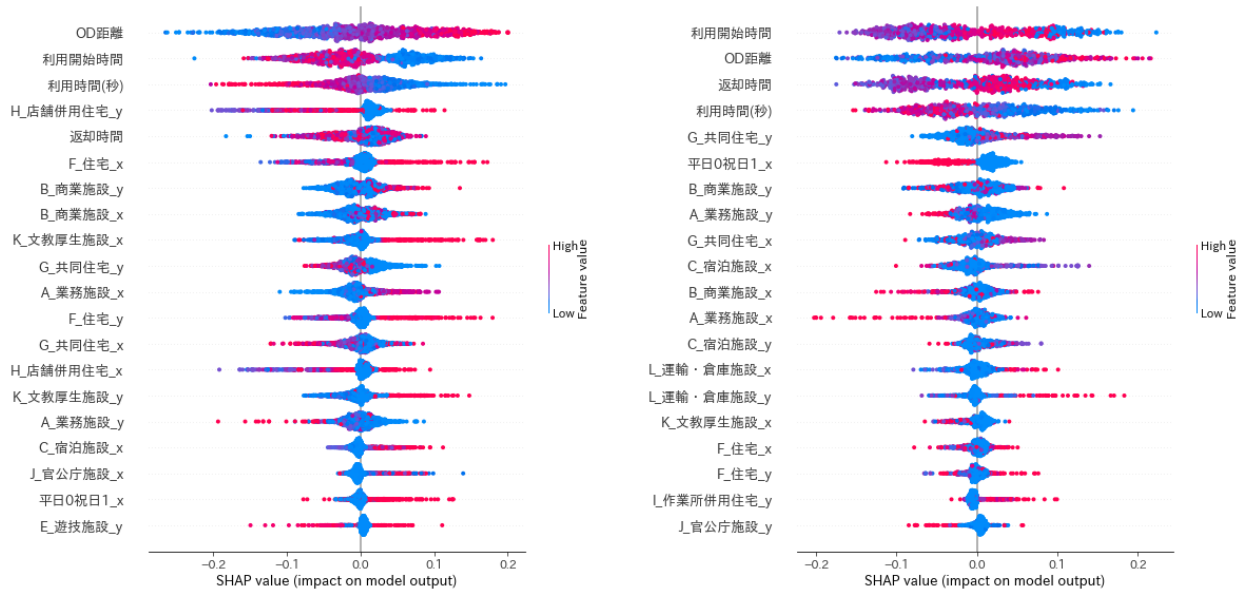


図-5 通勤目的の Shapley Value XGBoost (左：月額会員 右：1回会員)

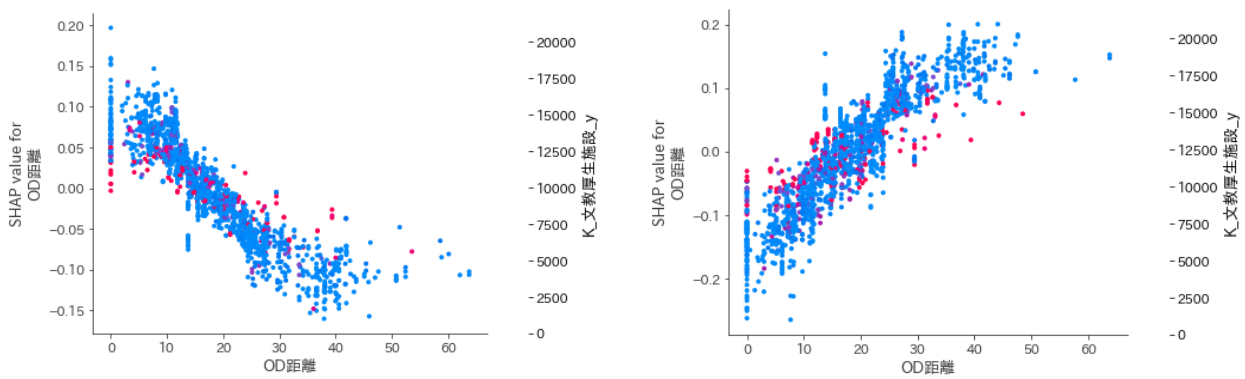


図-6 通勤目的の Shapley Value と OD 距離の関係図

(左：ランダムフォレスト 月額会員 右：XGBoost 月額会員)

## 6. まとめ

本研究では、ランダムフォレスト、XGBoost、ロジスティック回帰を用いてアンケート利用データから全トリップの利用目的の推定を行った。その結果を以下に示す。

- ・ランダムフォレスト、XGBoost とともに通勤目的、通勤以外目的の予測精度が高く、概ね推定精度が得られたといえる。
- ・ランダムフォレスト、XGBoost を用いて、2019 年から 2022 年までの全利用データに対し、利用目的を推定することにより、月額会員は総利用回数は増えているが、2019 年度-2022 年度を通しての特徴はみられなかった。1 回会員は通勤以外目的で利用する傾向がみられた。
- ・SHAP を用いることでそれぞれの予測モデルの特徴量をつかむことができ、OD 距離、返却時間、利用時間(秒)、利用開始時間が重要な説明変数であることを示した。

課題としては、本研究では通勤とそれ以外に利用目的を分けてモデルを作成したため、それ以外の中に分類される観光やショッピング等の行動分析ができないこと、アンケートデータの「ポロクルの利用目的は？(2 つまで選択可能)」という項目において、2 つ回答している利用者データの分析が不十分であることが挙げられる。今後 COVID-19 の感染拡大が収束し、通勤目的以外でポロクルが多く使われるようになったときに再び細かく利用項目ごとに予測モデルを作成し分析していきたい。

謝辞：本研究では、NPO 法人ポロクルからの貴重なデータを頂きました。ここに記して感謝の意を表します。

## 参考文献

- 1) 国土交通省：自転車活用推進計画, <https://www.mlit.go.jp/road/bicycleuse/pdf/2plan.pdf>, (2023/02/24 閲覧)
- 2) 国土交通省：第 4 回 シェアサイクルの在り方検討委員会, <https://www.mlit.go.jp/road/ir/ir-council/sharecycle/pdf04/02.pdf>, (2023/02/24 閲覧)
- 3) ポロクルホームページ：<https://porocle.jp/>, (2023/02/24 閲覧)
- 4) 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・講演集, 2020
- 5) 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
- 6) 大井啓史, 野崎脩人, 坂本信, 浅田拓海, 有村幹治：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, 第 63 回土木計画学研究発表会, 2021
- 7) 波部斉：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
- 8) 板橋将之, 本田あおい, 大北剛：SHAP 値や重要度を用いたモデル解釈性:包除積分ネットワークと XGBoost の比較, 火の国情報シンポジウム 2020 情報処理学会九州支部, 2020
- 9) ロジスティック回帰分析と傾向スコア(propensityscore) 解析, 天理医学紀要 2016 年 19 巻 2 号 p.71-79, 2016

## ESTIMATING TRIP PURPOSES OF A SHARED BICYCLE SYSTEM APPLYING A MACHINE LEARNING

Hyoga TANDO, Hirofumi OI,  
Takumi ASADA and Mikiharu ARIMURA