

Explainable AI を活用した 機械学習型交通行動モデルの解釈可能性

石嶋 悠嗣¹・柳沼 秀樹²・寺部 慎太郎³・海野 遥香⁴

¹ 学生非会員 東京理科大学大学院 理工学研究科土木工学専攻修士課程

(〒278-8510 千葉県野田市山崎 2641) E-mail: 7621502@ed.tus.ac.jp

² 正会員 東京理科大学 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)

E-mail: yaginuma@rs.tus.ac.jp

³ 正会員 東京理科大学 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)

E-mail: terabe@rs.tus.ac.jp

⁴ 正会員 東京理科大学 理工学部土木工学科 (〒278-8510 千葉県野田市山崎2641)

E-mail: unoharuka@rs.tus.ac.jp

離散選択モデルは、交通計画分野をはじめ様々なシーンで活用されており、ランダム効用理論を下敷とした人間の選択行動を理解する有用な手法である。しかしながら、離散選択モデルは理論性と解釈性を重視した手法であり、多数の選択肢を持つ場合では精度が低下するなどの問題があった。本研究では、機械学習を援用した予測精度の向上とモデルの予測根拠を示す解釈指標を用いることによって、「精度」と「解釈性」の両面で優れた予測手法の構築を試みる。具体的には、ニューラルネットワークを活用した交通手段選択モデルを構築し、PD(Partial Dependence)やSHAP(Shapley Additive exPlanations)などのXAIによる解釈指標を適用することで、解釈可能かつ予測精度が優れたモデル構築を図った。その結果、従来モデルよりも予測精度が高く、解釈性においてもより詳細な評価が可能になり、施策評価の高度化に寄与することが期待される。

Key Words: Discrete Choice Model, Neural Network, Explainable Artificial Intelligence(XAI)

1. はじめに

交通分野において、離散選択モデルはさまざまな場面で使用されてきた。目的地、交通手段、経路など、人間の選択行動原理を理解し予測する上では有用な手法の一つである。また、近年では離散選択モデル以外にも分析手法が発達してきており、特に機械学習による予測・分析は様々な場面で用いられるようになってきている。しかし、交通分野では画像解析や物体検知において機械学習の導入が進んできているものの、交通需要予測などの選択行動を予測する分野においては機械学習を用いた分析例は少ない。これは交通行動予測においては「解釈性」が重要視されるため、離散選択モデルのように人間が理解できるモデルでなければ扱いが難しいという点が原因だと考えられる。たとえ予測精度が良くても人がなぜその選択をしたのかを説明できなければ、政策評価やシナリオの予測ができないため実務で用いることが困難である。そこで本研究では以上の課題を踏まえ、機械学習に

よる精度向上とそのモデルの予測根拠を示す解釈指標を用いることによって、「精度」と「解釈性」の両面で優れた予測手法の構築を試みる。具体的には、交通手段選択モデルをニューラルネットワークにより構築し機械学習モデルの解釈性、いわゆるXAI(eXplainable AI: XAI)における4つの解釈ツール(PI,PD,CPD,SHAP)を用いてモデルの解釈性を検討した。それにより、従前の離散選択モデルよりも精度と解釈性の両面で優れたモデルを構築し、施策評価の高度化を目指す。

2. 離散選択モデルとニューラルネットワーク

(1) 離散選択モデル

交通行動分析において離散選択モデルは、人の選択行動を理解し予測する上で非常に強力なツールとして使用されている。特にこれまでの離散選択モデルでは、効用関数が線形のものも多く用いられている。これは効用関

数を線形に仮定することによって、非線形なものよりも取り扱いが容易になり、解釈がしやすいという利点がある。また、線形化は時間や費用に対して弾力性を把握することにも繋がり政策評価への解釈が可能である点や、計算コストの軽減などの利点がある。そして、効用関数の線形化は、尤度関数の凸性が保証され Newton-Raphson 法等によりパラメータを容易に推定できるという利点も挙げられる。これらの理由から離散選択モデルの効用関数は線形に仮定されることが多く、様々な場面で用いられてきた。しかし、このような離散選択モデルにもいくつかの問題がある。

1 つ目の問題点は人間の選択行動は必ずしも線形的ではないという点である。交通手段選択モデルを例に取って考える。効用関数に「移動時間」という変数を組み込みパラメータの推定を行うと、パラメータの符号は負になると考えられ、移動時間が増えれば増えるほど選択確率は小さくなると解釈できる。しかし、これは実際の行動に即していないと考えられる。例えば、徒歩を選択する場合、移動時間が 10 分程度であれば選択されやすく、10 分より長くなると選択されにくくなると考えられる。つまり、移動時間と効用関数の関係は非線形であり、離散選択モデルにおける効用関数の線形化は不適切であると考えられる。移動時間と選択確率の関係には単調増加・減少だけではなく変曲点があるはずであり、この現象は線形的な効用関数では表現することができない。

2 つ目は予測精度の低さである。上記でも述べたように離散選択モデルでは効用関数を線形に仮定し解釈性を重視しているため、予測精度の低さが課題として挙げられている。交通手段選択モデルのように選択肢が少ない場合は問題ないが、目的地選択のように選択肢が多くなると精度は低下してしまうという問題がある。

(2) ニューラルネットワーク

本研究ではニューラルネットワークを用いて交通手段選択モデルを構築する。ニューラルネットワークをはじめとする機械学習は近年目覚ましい発展を遂げており、モデル構造や予測精度は従来の機械学習よりも優れたものが開発され続けている。近年では交通の分野においても機械学習は利用されており、CCTV カメラの映像を用いて交通量観測をするなど画像認識の分野では使用されつつある。しかし、人間の行動原理・特性を予測する研究においては、機械学習が用いられることはほとんどない。行動モデルにおいては「解釈性」が重要視されるため、機械学習のようなブラックボックスのモデルでは扱いきれないことが原因だと考えられる。離散選択モデルのように線形ではなく、非線形に考えることで現象をより複雑に記述し予測精度を高めることが可能になるものの、その複雑さから解釈することは困難であるという問題が

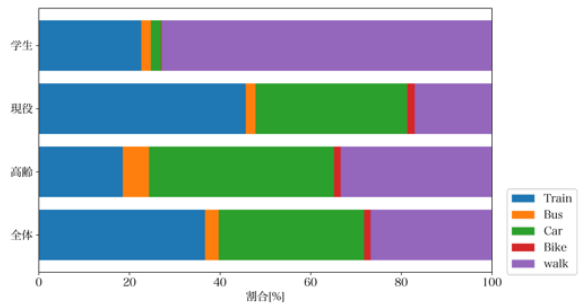


図-1 東京 PT 交通手段の基礎集計

ある。

3. データの概要と基礎集計分析

(1) データの概要

本研究では、H30 東京パーソントリップ調査データ（以下、東京 PT データ）を用いた。東京 PT データは個人の平日 1 日のトリップデータを収集している。東京を中心とする半径約 80km 圏域を対象とした調査であり、平成 30 年の 9-11 月に行われた。本研究では、この東京 PT データを用いて交通手段選択モデルを構築する。

(2) 基礎集計分析

交通手段分担率を図 1 に示す。全体の分担率を見ると鉄道、車、徒歩が主要な交通手段であることがわかる。また、トリップを学生(19 歳まで)、現役(20-64 歳または 65 歳以上の就業者)、高齢(65 歳以上)に分けて見ると、学生は、鉄道と徒歩がほとんどである。現役世代は、車の利用が多くなっており、鉄道と自動車の割合が多くなっている。高齢の人々に関しても自動車を利用する人が多く、鉄道を利用する人の割合は小さくなっている傾向があることがわかる。

4. モデル構造

(1) 離散選択モデル

離散選択モデルは、単に個人の交通行動を記述するのではなく、個人が「利用可能な選択群の中から最も望ましい選択肢を選ぶ」といったルールに基づいて行動することを仮定してモデル化するものであり、選択肢 n を選ぶ確率は式(1)で表すことができる。

$$P(n) = \frac{\exp(V_n)}{\sum_l \exp(V_l)} \quad (1)$$

(2) ニューラルネットワーク(NN)

NN は脳の神経細胞（ニューロン）を模倣したものである。1つのニューロンには、ニューロン同士のつながりの強さを表す重みと、受け取った値を次層にどのように出力するかを決める活性化関数で構成され、これらを多数結合させたものが NN である。大きく分けると外部からの情報を受け取る入力層、上位の中間層または出力層に情報を送る中間層、外部に情報を送る出力層の 3 つで構成される。

本研究では、一般的なネットワークではなく離散選択モデルにおける選択肢集合を加味させるために「スキップコネクション」を用いてネットワークを表現した。スキップコネクションとは、主に画像解析の機械学習で用いられる手法である。従来まで CNN などのモデルは層を深くすることで精度向上を図っていたが、ある一定の層を超えると精度は向上せず勾配消失問題や劣化問題が起きてしまうという問題が生じていた。そこでこれらの問題を解決するために Kaiming He がネットワークをスキップさせる 'Res-Net' を考案し、これらの理論を参考にニューラルネットワークにスキップコネクションを導入した。具体的には図 2 のようなモデル構造である。入力層には説明変数に加え各個人を選択肢集合を使用し、中間層に移る前に特徴量と選択肢集合を分け、特徴量のみを中間層のネットワークに通す。最後に、中間層の出力値と選択肢集合を掛け合わせることで不要な選択肢の重みを軽減させることによって選択肢集合を加味した。しかし、本研究で用いたネットワーク構造は正確に選択肢集合を反映させたものでないという点に留意したい。選択肢集合を掛け合わせている部分はあくまで中間層であり、中間層と出力層は全結合によって結ばれているため、出力層においては全ての選択肢が対象であり、選択肢集合が 0 の選択肢についても確率の計算がされてしまう。離散選択モデルでは選択肢集合が 0 の選択肢に関しては計算される確率も 0 になるため、出力値の計算方法は離散選択モデルと本研究のニューラルネットワークで異なる点である。

5. 解釈性の評価手法

(1) 機械学習の解釈性

近年、機械学習において解釈性が重要視されている。従来までの機械学習では、ブラックボックスであったため、特徴量がどれだけ予測値に影響を与えているのか、どうしてそのモデルがこのような出力結果を算出したのかを解釈することは困難であった。そのため、モデルを開発する際にもデバック処理の難しさや、精度向上に向けた特徴量エンジニアリングのアプローチへの困難さが課題として挙げられていた。また、モデルの不透明性は

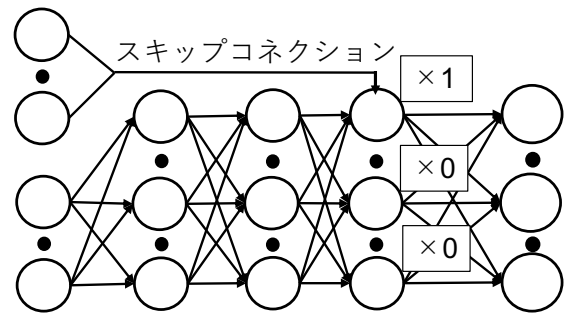


図-2 スキップコネクションの概要図

モデルの扱いやすさに影響を与えるだけでなく、信頼性にも影響を与える可能性がある。解釈ができないモデルは業務を行う上でも信頼が得られにくく、理解が必要とされる場面においては不向きであった。しかし、近年これらの問題を解消できるような解釈可能ツールが多く開発されてきている。そして本研究では、ニューラルネットワークを用いた交通手段選択モデルに4つの解釈手法を適用し、モデルの解釈性・理解度を示す。具体的には、Permutation Importance(PI)、Partial Dependence(PD)、Conditional Partial Dependence(CPD)、SHapley Additive exPlanations(SHAP)の4つの解釈ツールを利用して機械学習モデルの解釈を試みる。従来までの機械学習による予測では、学習過程がブラックボックスであるためどの特徴量がどれだけ学習に寄与したのか、どのように目的変数に寄与したのか、なぜモデルがそのような予測値を算出したのかを判断することが困難であった。仮にランダムフォレストのようなアルゴリズムであれば、Gini係数を用いて重要度などの指標を計算することは可能であるが、NNのように複雑なアルゴリズムには適応することができない。しかし、本研究で用いる上記の4つの指標はアルゴリズムに依存せずに特徴量の重要度や予測値に対する影響を計算することが可能である。

(2) Permutation Importance (PI)

PIは2018年にFisherらによって考案された指標であり、ある特徴量がどれだけモデルの予測精度向上に寄与しているのかという特徴量重要度を計算するための指標である。PIを用いることによって、どの特徴量がモデルの予測に強く影響しているのかを判断することができる。この指標を用いることによりモデルのデバックや効率的な特徴量選択が可能になる。このようにモデルに使用した特徴量の重要度を知ることはモデルの振る舞いを確認し、モデルの精度を高める上で非常に重要な役割を果たす。

PIは特徴量の値をシャッフルするし、その特徴量の情報が使えない状態を作り出すことで算出できる。手順としては学習させたモデルを用意し、使用した特徴量のうち1つをシャッフルして予測を行う。最初に学習させた予測誤差と特徴量の1つをシャッフルして計算された予

測誤差の差分がその特徴量の重要度になる。この手順を他の特徴量に対しても行うことで、全ての特徴量重要度を比較することができる。しかし、PIはあくまでモデルを構成する上で重要な変数を表す指標であるため、重要度から因果的な解釈を行うことは危険である。各特徴量がどのようにモデルの予測に寄与しているのかを判断することはできない

(3) Partial Dependence (PD)

PDは、2001年にJ.H.Freidmanにより考案されたものである。1つまたは2つの特徴量がモデルの予測結果に与える周辺効果を計算することで、特徴量と予測値の関係を知らることが可能である。ある特徴量と目的変数が線形もしくは非線形の関係にあるのか、比例・反比例の関係にあるのかなどを表現することができる。考え方としては、他の特徴量を固定し知りたい特徴量のみを動かし、各インスタンスの予測値を平均したものがPDとなる。これは数学における周辺化と呼ばれる操作に相当し、以下の式のように表現できる。

$$\hat{f}_s(x_s) = E_c[\hat{f}(x_s, x_c)] = \int \hat{f}(x_s, x_c)p(x_c)dx_c \quad (2)$$

x_s はPDを計算する対象となる特徴量、 x_c はその他の特徴量、 \hat{f} は学習済みのモデルである。そしてこの式を元に以下の式を計算することによりPDを計算することができる。

$$\bar{f}_s(x_s) = \frac{1}{N} \sum_i \hat{f}(x_s, x_c^{(i)}) \quad (3)$$

(4) Conditional Partial Dependence (CPD)

CPDは2019年にZhaoらによって考案された解釈指標であり、PDで考慮できなかった「異質性」を反映した指標である。PDでは特徴量と予測値の関係を把握するために各インスタンスの平均を用いて計算したが、実際には特徴量と予測値の関係はインスタンスごとに異なることが考えられる。例えば、交通手段と移動時間の関係を考えても、各個人で時間に対する考え方は異なることが予想される。ある学生は比較的時間に余裕があるため移動時間が増えても選択確率にはそこまで影響を与えず、逆にある社会人は時間短縮を優先するため、一定の時間を超えると選択確率が下がる可能性が考えられる。このような場合にPDを用いて交通手段と移動時間の関係を見てしまうと、「平均的」な関係を示すPDでは異質性に気づくことができない。

そこでPDのように全体の平均を取るのではなく、各グループごとの関係に焦点を当てた方法がCPDである。

$$f_s(x_s)|_{x_g=v} = E_{xc}f(x_s, x_c, x_g = v) \quad (4)$$

CPDは上記のような式で表現され、あるカテゴリ変数 g が v を満たすインスタンスのみでPDを計算することによって算出される。

(5) SHapley Additive exPlanations (SHAP)

SHAPは機械学習において、モデルがなぜそのような予測値を出したのかを説明するための手法である。SHAPを用いることによってPIのようなモデル全体に対する解釈に加えて、個人に焦点を当てた解釈も可能であり、マクロとミクロの両方の解釈が可能である。算出方法は協力ゲーム理論のシャープレイ値を応用している。シャープレイ値をそのまま算出するには変数の数が増えると組み合わせが増えてしまい、計算量が膨大になってしまうため、計算方法を工夫することによって機械学習で扱えるようにしたものがSHAPである。これを計算することによって一つ一つの特徴量がモデル予測値に与える貢献度を計算することが可能になり、以下の式のように表現される。

$$\phi_i = \frac{1}{|F|!} \sum_{S \subseteq F \setminus \{i\}} |S|!(|F| - |S| - 1)! [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (5)$$

F は特徴量の数、 S はモデルで使用されている特徴量の部分集合、 x は説明されるインスタンスの特徴量ベクトルである。

6. モデルの結果と解釈性の考察

(1) 推定結果

東京PTデータを用いて、鉄道、バス、自動車、バイク、徒歩の5つの交通手段予測モデルを構築した。推定結果を表1に示す。離散選択モデルでの推定に関しては、ほぼ全ての変数が有意となりパラメータの符号も良好であった。また、尤度比も0.2を超えておりの中率も約7割であることから優れたモデルであると言える。

ニューラルネットワークに関しては表2の変数を入力層に入れ、各交通手段の選択確率を算出した。また、前述したように選択肢集合に関してはスキップコネクションを用いて出力層の前に掛け合わせることで表現した。その結果、ニューラルネットワークの的中率は0.8となり、離散選択モデルよりも約10%的中率を向上させることができた。

(2) 解釈性指標の考察

次に、PI、PD、CPD、SHAPをニューラルネットワー

表-1 推定結果

変数	パラメータ	t値
定数項(train)	1.61	3.94 ***
定数項(bus)	0.354	0.800
定数項(car)	4.02	9.15 ***
定数項(bike)	4.64	10.2 ***
費用	-0.167	-3.39 ***
免許保有ダミー	2.67	3.94 ***
短距離ダミー	1.19	5.56 ***
アクセス距離	0.436	7.36 ***
時間	-0.451	-8.45 ***
都心ダミー	-0.990	-5.21 ***
サンプル数		1000
初期尤度		-1608
最終尤度		-888
修正済み尤度比		0.441
的中率		0.676

「***」:1%有意, 「**」:5%有意, 「*」:10%有意

クに適用し, 機械学習に対して解釈を行っていく.

まずはPIの結果を図3に示す. この図からモデルを構成する上で重要な変数は徒歩の選択肢集合, 移動時間, 車の費用だとわかる. 徒歩の選択肢集合に関しては, 移動距離に基づいて算出しており, ある一定の距離以上は徒歩を選択しないと仮定している. そのため, 選択肢集合自体よりも「移動距離」がモデルに大きく影響していると考えられる. 同様に, 移動時間に関しても距離に比例するものであるため, 重要度が大きくなったと考えられる. 車の費用に関しては距離の影響も考えられるが, もう一つの影響として全体の分担率が大きく影響していると考えられる. 今回のデータでは鉄道, 自動車, 徒歩の分担率がそれぞれ約3割ずつである. そして, 徒歩ト

表-2 交通手段選択モデルで使した変数

使用した変数
性別, 年齢, 職業, 収入, 移動目的, 出発ゾーン, 到着ゾーン, 滞在時間, 出発施設, 到着施設, 各移動時間, 各移動料金, アクセス・イグレス距離, 選択肢集合

リップの大部分は学生でありそれ以外のトリップは15-64歳の現役世代が占めている. よって, これらの人々にとってはトリップの大半が鉄道か自動車の二択を選ぶ状況にあると考えられる. そのような状況になったときに, 鉄道は誰もが手軽に安価で利用できる一方で, 自動車はガソリン代や所有コスト, 駐車場料金などがあるため移動手段の判断に大きく影響すると考えられる. そのため, 鉄道コストよりも特徴量の重要度としては大きくなり, モデル全体としても大きくなったと考えられる.

次にPDの結果を図4-8に示す. 各交通手段によって縦軸の選択確率が異なる点に注意した上で考察していく. 今回の分析においては「移動時間」と各交通手段の選択確率の関係性を可視化した. 特徴量重要度でも上位に来ており, かつ連続量であるという点から選定した. また, 移動時間以外でも連続量であればPDを計算することは可能であり, 移動時間の他にも滞在時間や年齢なども選択確率の関係性を可視化することが可能である. 鉄道での移動時間と選択確率の関係性は図4に示す. 40分までの移動に関しては線形的に増えておりそれ以上の移動になると選択確率は変化しないことがわかる. これは実際の選択行動に即しており, 長距離の移動に関しては鉄道を使用しやすくなるという解釈が可能である. バイクと徒歩に関しては移動時間がある一定の値を超えると選

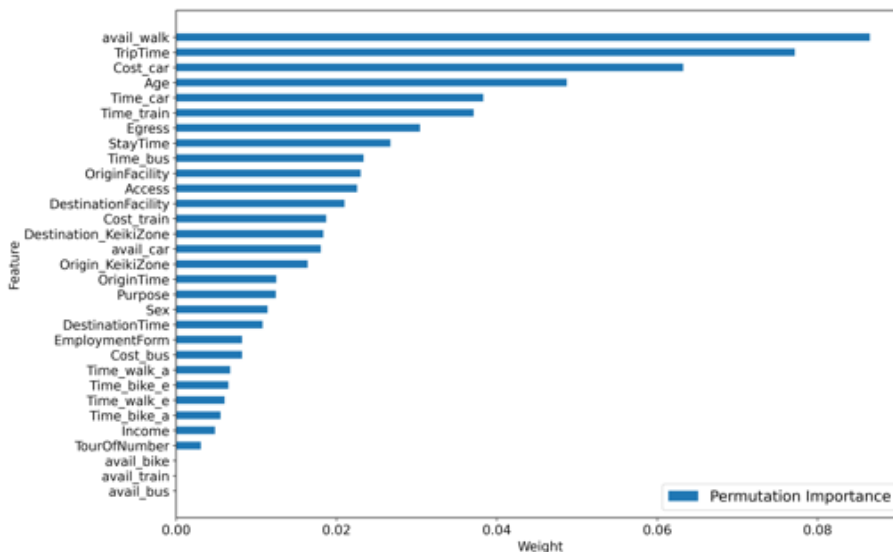


図-3 Permutation Importance

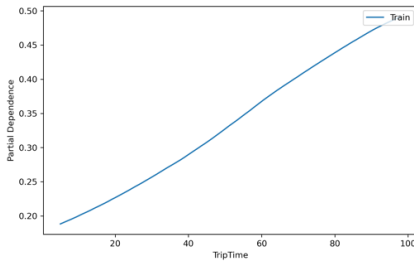


図-4 鉄道の PD

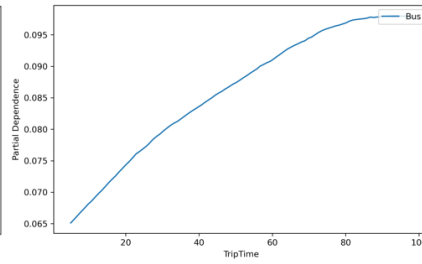


図-5 バスの PD

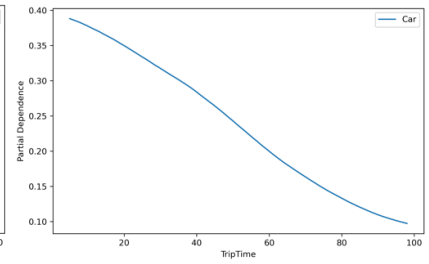


図-6 自動車の PD

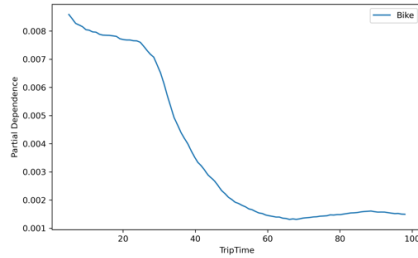


図-7 バイクの PD

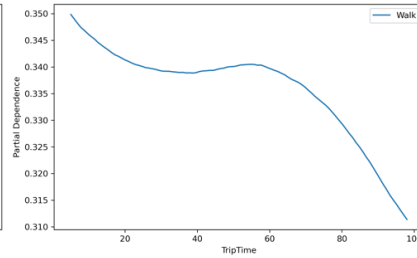


図-8 徒歩の PD

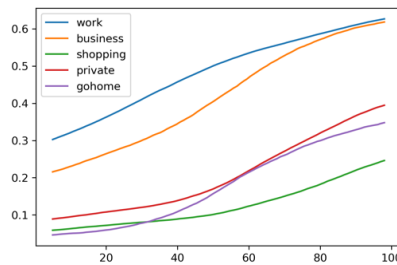


図-9 鉄道の CPD

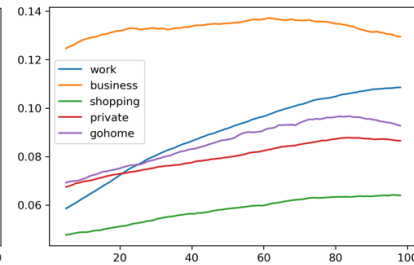


図-10 バスの CPD

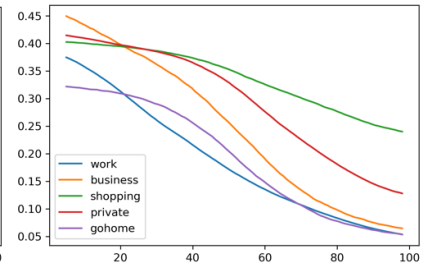


図-11 自動車の CPD

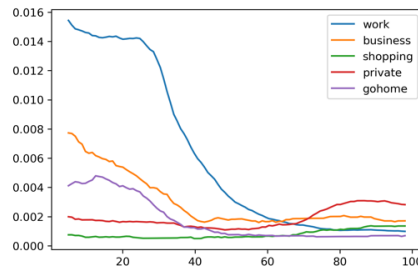


図-12 バイクの CPD

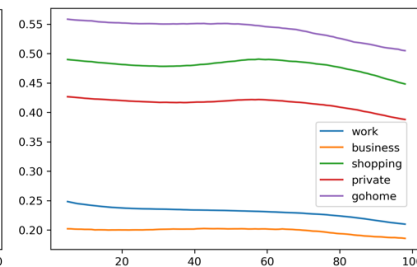


図-13 徒歩の CPD

択確率が上昇する結果となった。これに関しては PD の特性でもある範囲外の値に関しては信頼性の低い結果になるという影響が反映されていると考えられる。今回の分析ではデータ処理の段階で外れ値と思われるトリップは除外しており、バイクと徒歩に関しては長距離の移動は存在しない状態で推定している。そのため、バイクと徒歩に関しては長距離の PD の表現性が不安定になり、実際の選択行動とは反する結果になったと考えられる。

CPD の結果を図 9-13 に示す。今回は移動目的でグループ分けを行い、「通勤」「業務」「買い物」「私事」「帰宅」の 5 つに分類した。CPD も PD と同様の関係性を見ることができ、大きな選択確率の変動があった鉄道について考察する。鉄道に関しては通勤と業務において線形の関係が確認でき、この 2 つのトリップに関しては移動時間が長くなればなるほど選択確率が大きくなるこ

とがわかる。それ以外のトリップに関しては 40 分以上の移動では大きな変化がなく選択確率に与える影響は小さいと考えられる。

最後に SHAP での解釈性結果を示す。SHAP の中には複数の解釈手法が存在しており、各インスタンスの予測結果を解釈するミクロな手法から、粒度を荒くし集計を行うことで PI や PD のようなマクロな解釈手法としても使用することができる。図 14 は目的変数に対する変数の寄与度を可視化したものである。PI と似ているものの「何を持って重要な変数とするか」の定義の部分が異なっている。PI ではモデルの性能低下に基づいて重要度を計算しており、SHAP に関しては特徴量の帰属の大きさ、つまりどれだけ目的変数に影響を与えているかという点に基づいて計算しており、この点において特徴量の重要度が異なっている。

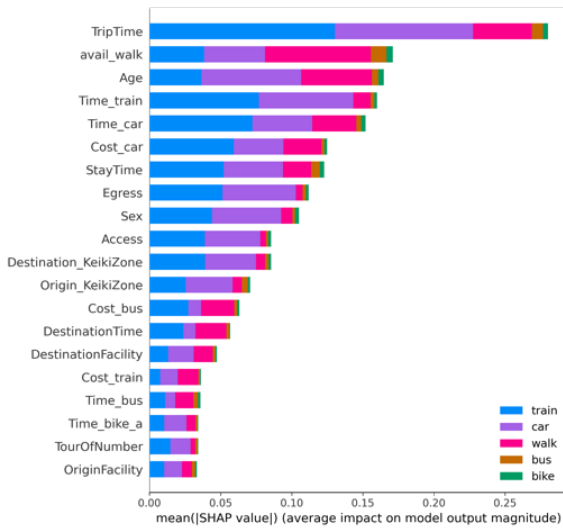


図-13全体の summary_plot

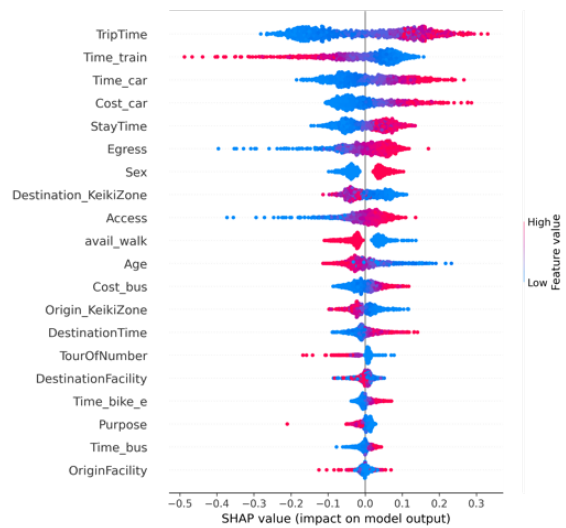


図-14鉄道の summary_plot

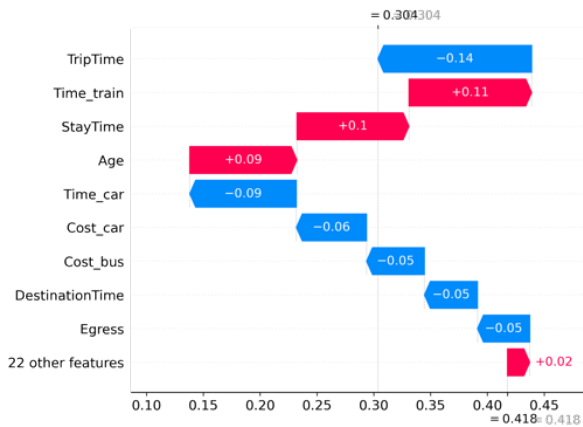


図-15鉄道の waterfall_plot

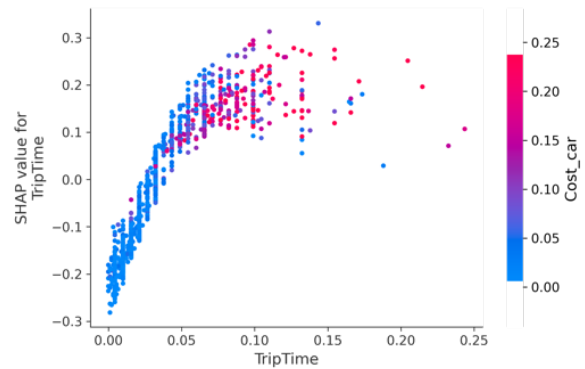


図-16鉄道の dependence_plot

図 15 は図 14 と同様に特徴量重要度を示しており、それをプロットしたものである。今回は交通手段が鉄道における特徴量重要度のみを示しているため、変数の順序は変化しているが大きな変化はないと言えるだろう。また、プロットの色はその特徴量の高低を表しており、例えば今回の結果では、移動時間が大きければ SHAP 値も大きくなりやすいといった性質を知ることができる。特に特徴量重要度で上位にある鉄道と自動車の移動時間の大きさに関しては対局的であり、トレードオフの関係ということもわかる。

図 16 はモデルがどのような経緯で選択確率を算出したのかを可視化したものである。今回はある個人の鉄道における選択確率の変動を示したものである。各変数が選択肢に与える影響を得ることができ、離散選択モデルでのパラメータと同様の解釈をすることが可能である。さらに、離散選択モデルではデータ全体のパラメータしか見ることができなかったものの、SHAP を用いることで個人のパラメータを得ることができマイクロレベルでの解釈が可能になった。

図 17 は図 16 の値を全てのサンプルにおいて集計し、

マクロレベルで各変数と SHAP 値の関係性を可視化したものである。横軸に移動時間、縦軸に SHAP 値をとったものであり、鉄道のみに対して集計したものである。移動時間が増えるほど予測値にプラスに貢献しやすくなる傾向がわかる。さらにカラーバーに表示する変数を指定することにより、特徴量同士の交互作用も可視化することができる。今回はカラーバーに自動車の費用を設定した。これにより、移動時間が長い人の中でもさらに自動車の費用が高い人は SHAP 値が大きくなり、鉄道を選択しやすくなると解釈することができる。

5. おわりに

今回の分析では、交通手段選択モデルにおいて離散選択モデルとニューラルネットワークの比較を行い、精度と解釈性について比較を行った。まず、精度については離散選択モデルでも十分精度はあったものの、ニューラルネットワークを用いることによってさらに精度向上が確認され、主に 3 つの点が精度向上に影響したと考える。1 つ目は使用した変数である。離散選択モデルでは上手

く活用することができなかつたゾーン番号や発着施設を、ニューラルネットワークは容易に組み込むことができた。選択確率に影響を与える特徴量を増やしたことにより精度向上に繋げることができたと考える。2 つ目はモデルの非線形化である。離散選択モデルでは、効用関数を線形化することにより解釈性や計算コストの考慮していたが、ニューラルネットワークでは非線形にすることで複雑な現象を表現することが可能になった。3 つ目はスキップコネクションの導入である。離散選択モデルで使用する選択肢集合の概念をスキップコネクションを用いることによって考慮したことによって精度向上に寄与することができた。

解釈性に関しても様々な解釈指標を用いることによって、複雑な非線形モデルに対しても理解を得ることが可能になった。PI では、モデルの特徴量重要度を算出することで交通手段モデルにおいて必要な変数の把握が可能になった。これによりモデルを精度を決める変数が把握できるだけでなく、上位の変数がドメイン知識に沿っているかを確認することができモデルのデバックにも繋げることができる。PD, CPD では、各交通手段に対して移動時間と選択確率の関係を可視化した。実際の現象においては、各特徴量と目的変数の関係性は非線形であることが予想され、ニューラルネットワークのような複雑なモデルでも関係性を可視化することが可能になった。しかし、PD, CPD はあくまで特徴量と予測値の関係を見ているだけであり、明確な因果関係はないことは注意すべき点である。SHAP では、ミクロとマクロの視点から解釈することでモデルの振る舞いを可視化することができた。waterfall_plot のように個人ごとのパラメータを得ることや、summary_plot のように書く特徴量の影響をマクロに見ることも可能である。これらを用いることによって今までブラックボックスであった機械学習のような複雑なモデルにおいても解釈性を与えることが可能になった。そして、離散選択モデルと比較しても精度と解釈性の両面で優れたモデルを構築することができた。

今後の課題としてはモデル構造に関して 2 つの点が挙げられる。1 つ目は、スキップコネクションによる選択

肢集合だ。今回のモデル構造では正確に選択肢集合を反映できておらず、あくまで近似的に反映させている。本来であれば、出力層に必要なニューロンだけ softmax で変換する処理が必要である。しかし、本研究ではあくまでスキップコネクションを用いて近似しており、離散選択モデルの構造を完全に表現できていない。2 つ目は、弾力性についてだ。離散選択モデルでは線形に定式化することによって弾力性を見ることが可能であったが、ニューラルネットワークのような非線形モデルでは弾力性を上手く表現できない。スキップコネクションを利用して中間層を避け線形的にモデルに組み込めば弾力性を表現できるが、移動時間や費用などモデルに重要な変数を線形化してしまうとモデルの精度が下がってしまうという問題がある。この部分においてはまだ精度と解釈性がトレードオフの関係にあり目的に応じて適切に判断する必要がある。

参考文献

- 1) Xilei Zhao, Xiang Yan, Pascal Van Hentenryck: Modeling Heterogeneity in Mode-Switching Behavior Under a Mobility-on-Demand Transit System: An Interpretable Machine Learning Approach, arXiv:1902.02904v1, 2019
- 2) Xilei Zhao, Xiang Yan, Pascal Van Hentenryck: Modeling Stated Preference for Mobility-on-Demand Transit: A Comparison of Machine Learning and Logit Models, arXiv:1811.01315v2, 2019
- 3) Yafei Han, Christopher Zegras, Francisco Camara, Pereira, Moshe Ben-Akiva: A Neural-embedded Choice Model: TasteNet-MNL. Modeling Taste Heterogeneity with Flexibility and Interpretability, arXiv: 2002.00922v1, 2020
- 4) 中山晶一郎, 高山純一, 山下裕一郎: 非線形効用関数を持つロジットモデル: ニューラルネットワークを例に. 土木計画学研究・論文集, Vol. 20, no. 3, 2003

(Received ?)
(Accepted ?)

INTERPRETABILITY OF MACHINE-LEARNING BASED TRAVEL BEHAVIOUR MODELS USING XAI

Yushi ISHIJIMA, Hideki YAGINUMA, Shintaro TERABE and Haruka UNO

Discrete choice models have been used in various situations in the transportation field. However, the prediction accuracy of the discrete choice model has been a problem because it is an interpretive method. In this study, we attempted to construct a model that is superior in both accuracy and interpretability by using machine learning to improve accuracy and by using interpretability indices to understand the model.