



Explainable AIを活用した 機械学習モデルと離散選択モデルの比較

Comparison of machine learning model and discrete choice model using explainable AI

2022.9.25

東京理科大学

Tokyo University of Science Planning Lab

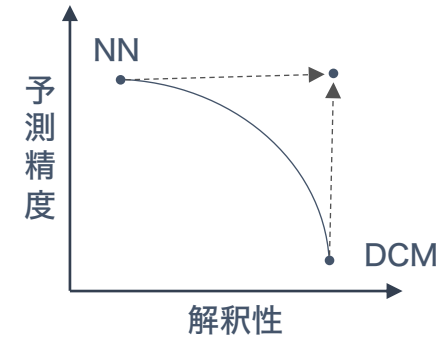




交通行動予測手法 Traffic behavior prediction method

離散選択モデル DCM

- 解釈性が高く施策評価手法として有効
- 「**解釈性**」「**理論性**」が重要視される分析手法
- 精度や処理速度などの**記述性能が求められる場面**での適用が困難
- Highly interpretable and effective as a policy evaluation method.
- An analysis method that emphasizes “Interpretability”, “theory”.
- Difficult to apply in situations where technical performance such as accuracy and processing speed is required.



機械学習 Machine learning

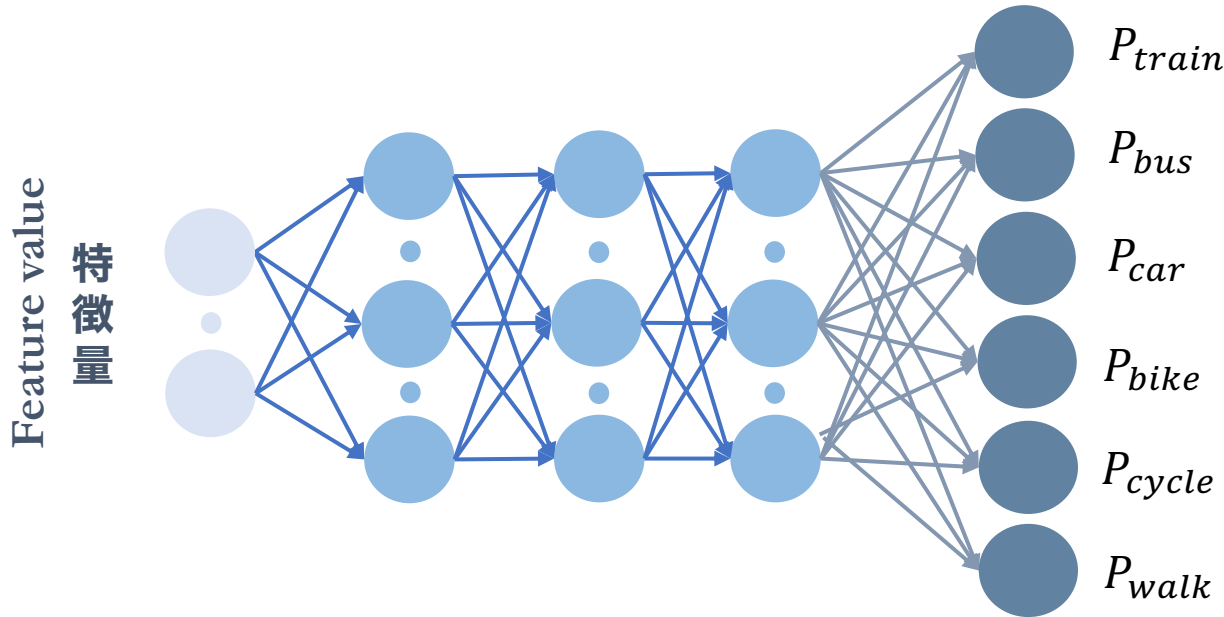
- 記述性能は高いが解釈性が乏しい
- 感度や予測根拠を示す**解釈指標(Explainable AI: XAI)**の発達
 - PI,PD,CPD,ICEなど
- ブラックボックスに対しても解釈が可能
- High technical performance but poor interpretability.
- Development of interpretive index (XAI) that shows sensitivity and predictive basis. →ex) PI, PD, CPD, ICE...
- Interpretation is also possible for black boxes.

機械学習と離散選択モデルの比較

Comparison of machine learning model and discrete choice model



手段選択モデル(NN) Mode choice model



- 中間層は4層
- ユニット数は順に 100,100,50,10
- 4 middle layers
- The number of units is 100, 100, 50, 10 in order

的中率 0.752
Accuracy 0.752

入力層 → H20東京PTデータ(特徴量)

Input layer → H20 Tokyo PT Data (Feature Value)

- 性別, 年齢, トリップ回数, トリップ順序, 出発ゾーン, 到着ゾーン, 移動目的, 出発時間, 到着時間, 滞在時間, **各移動時間**
- Sex, Age, Number of trips, Trip order, Departure zone, Arrival zone, Purpose of trip, Departure time, Arrival time, Stay time, **Each trip time**

出力層 → 選択確率

Output layer → Choice probability

- 鉄道, バス, 自動車, 二輪車, バイク, 徒歩
- Train, Bus, Automobile, motorcycle, bike, walk

しかし, 単純にNNで予測しただけでは**解釈性に乏しい**

➤ **XAI**を用いることにより**解釈性を与える**

Poor interpretability by simply predicting with NN → Giving interpretability by using **XAI**



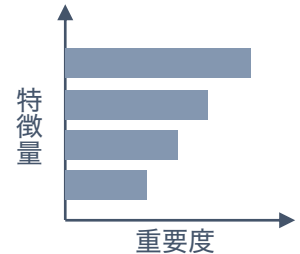
① Permutation Importance(PI)

- モデルにおける**特徴量重要度**を計算する指標

Calculating the feature importance in models.

- モデルの**精度を高める上で重要な特徴量**を判断する

Feature data is important to increase accuracy of models.



② Partial Dependence(PD)

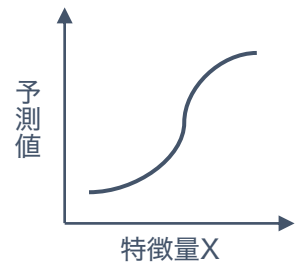
- 特徴量がモデルの予測値に**どのように影響を与えるのか**
特徴量と予測値の平均的な関係を計算する

Calculating the relationship between each feature and output.

- 特徴量と予測値が**比例か反比例か**、**線形か非線形か**を判断

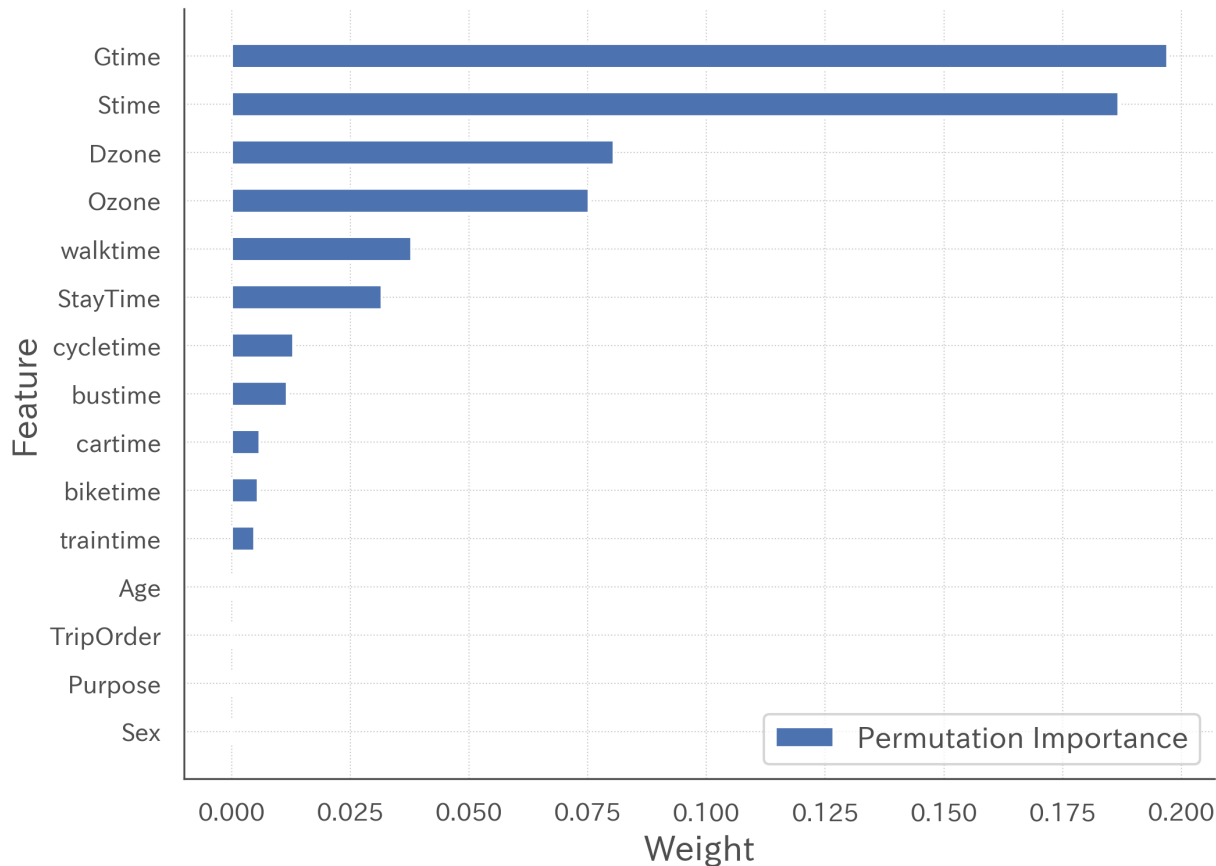
We can see whether the relationship is in proportion or in inverse proportion.

We can also see whether linear or nonlinear.





- NNの交通手段選択モデルにPIを適用 Apply PI to mode choice model
- 特徴量の重要度を計算 Calculating the feature importance in models.



Gtime: 到着時刻
Arrival time
Stime: 出発時刻
Departure time
Dzone: 到着ゾーン
destination zone
Ozone: 出発ゾーン
origin zone

- どの特徴量を優先的に入れるかを判断できる
We can judge which features should be put in models preferentially.
- 離散選択モデルでも変数の選択に用いることが可能
We can also set variables of DCM by looking at the bar graph.

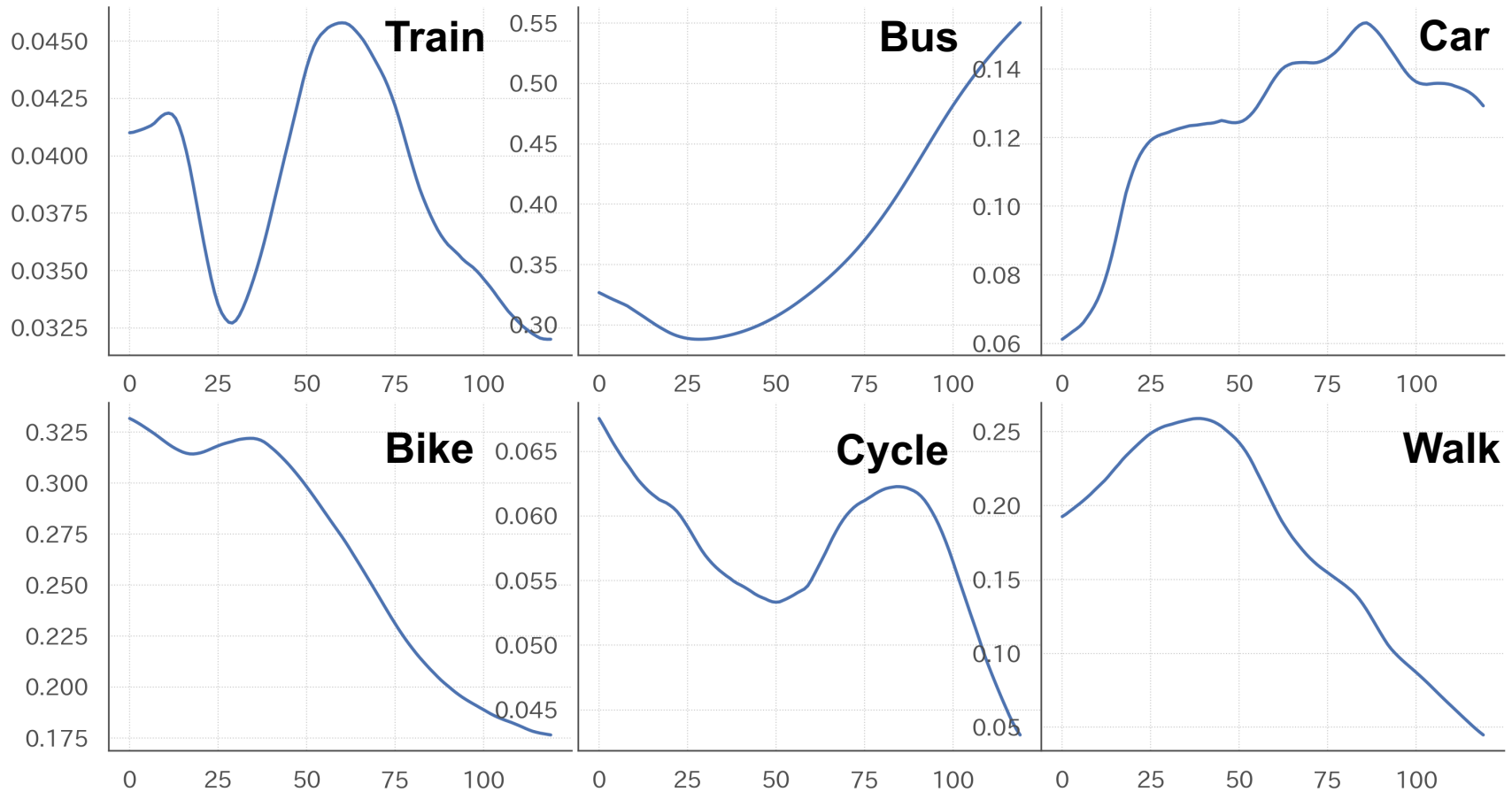


- PDを適用し，特徴量と予測値の関係性を分析

Calculate the relationship between features and predictions with PD

- ▶ **移動時間(横軸)と選択確率(縦軸)の関係**

horizontal axis is travel time. vertical axis is probability



- 非線形であるため，政策評価が難しい
Non-linear, making policy evaluation difficult

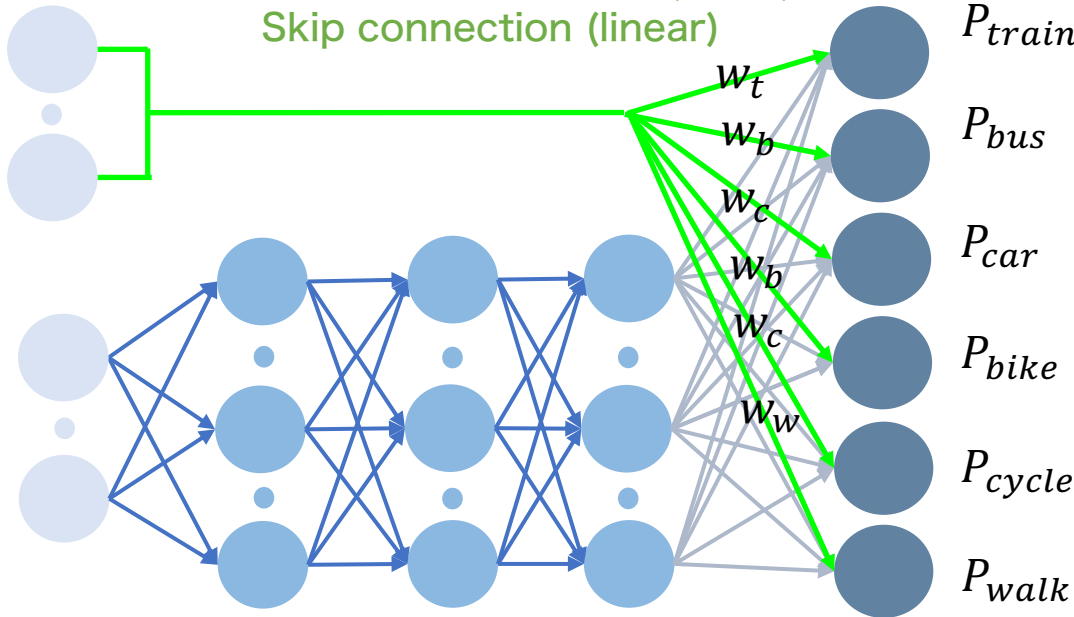


手段選択モデル(NN) Mode choice model

スキップコネクション(線形)
Skip connection (linear)

各移動時間
travel time

特徴量
feature



$$V_{ij} = V_{ij}^{MNL} + V_{ij}^{NN}$$

的中率 0.557
Accuracy 0.557

- $w_{train} = -2.63$
- $w_{bus} = -0.321$
- $w_{car} = -0.782$
- $w_{bike} = -0.678$
- $w_{cycle} = -0.327$
- $w_{walk} = -0.125$

入力層
input layer

→H20東京PTデータ(特徴量)
H20 Tokyo PT data (features)

出力層
output layer

→選択確率
Choice probability

- 性別, 年齢, トリップ回数, トリップ順序, 出発ゾーン, 到着ゾーン, 移動目的, 出発時間, 到着時間, 滞在時間
- Sex, age, trip number, trip order, Departure zone, Origin zone, purpose of trip, Departure time, Origin time, stay time

- 鉄道, バス, 自動車, 二輪車, バイク, 徒歩
- train, bus, car, bicycle, bike, walk
- これらの**選択確率**
- **These choice probabilities**



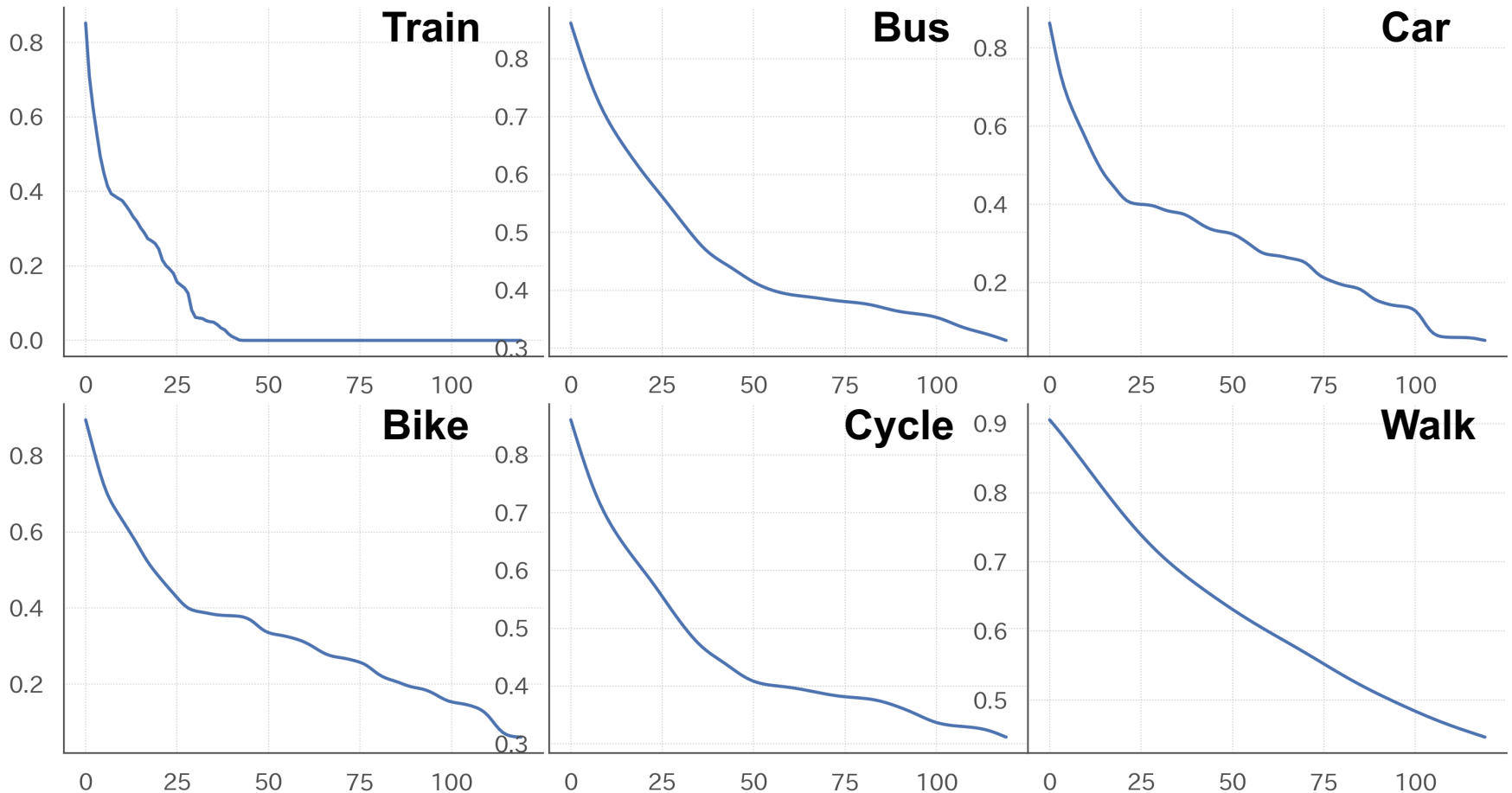
Partial Dependence(PD): NN+DCM

- PDを適用し，特徴量と予測値の関係性を分析

Calculate the relationship between features and predictions with PD

- **移動時間(横軸)と選択確率(縦軸)の関係**

horizontal axis is travel time. vertical axis is probability



- 選択確率が変曲点を持たない Selection probability has no inflection point



Partial Dependence(PD): NN+DCM

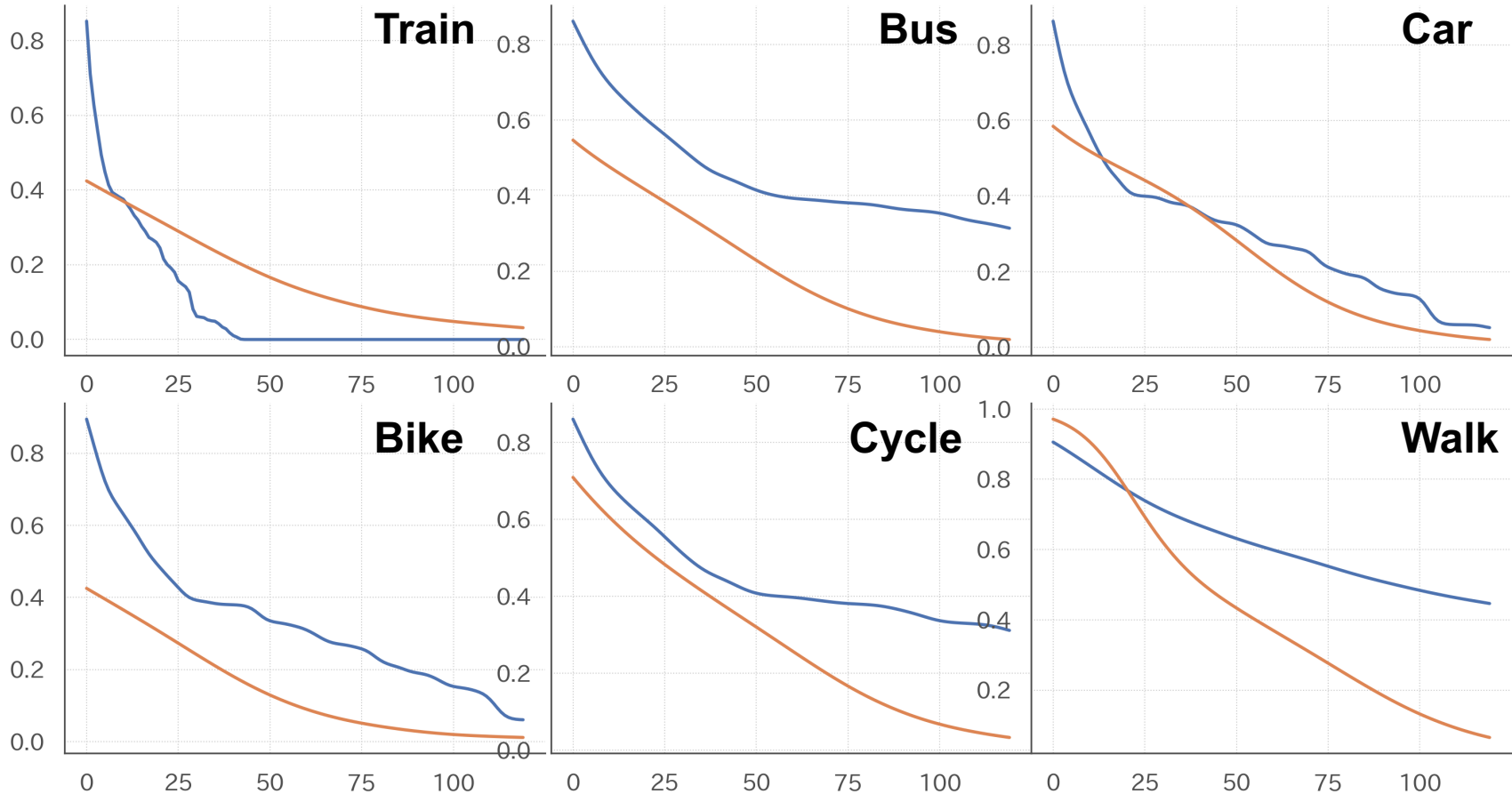


- PDを適用し，特徴量と予測値の関係性を分析

Calculate the relationship between features and predictions with PD

- **移動時間(横軸)と選択確率(縦軸)の関係**

horizontal axis is travel time. vertical axis is probability



- PDの欠点：存在しないデータに対して計算が不安定になる

Disadvantage of PD: Calculations are unstable for non-existent data



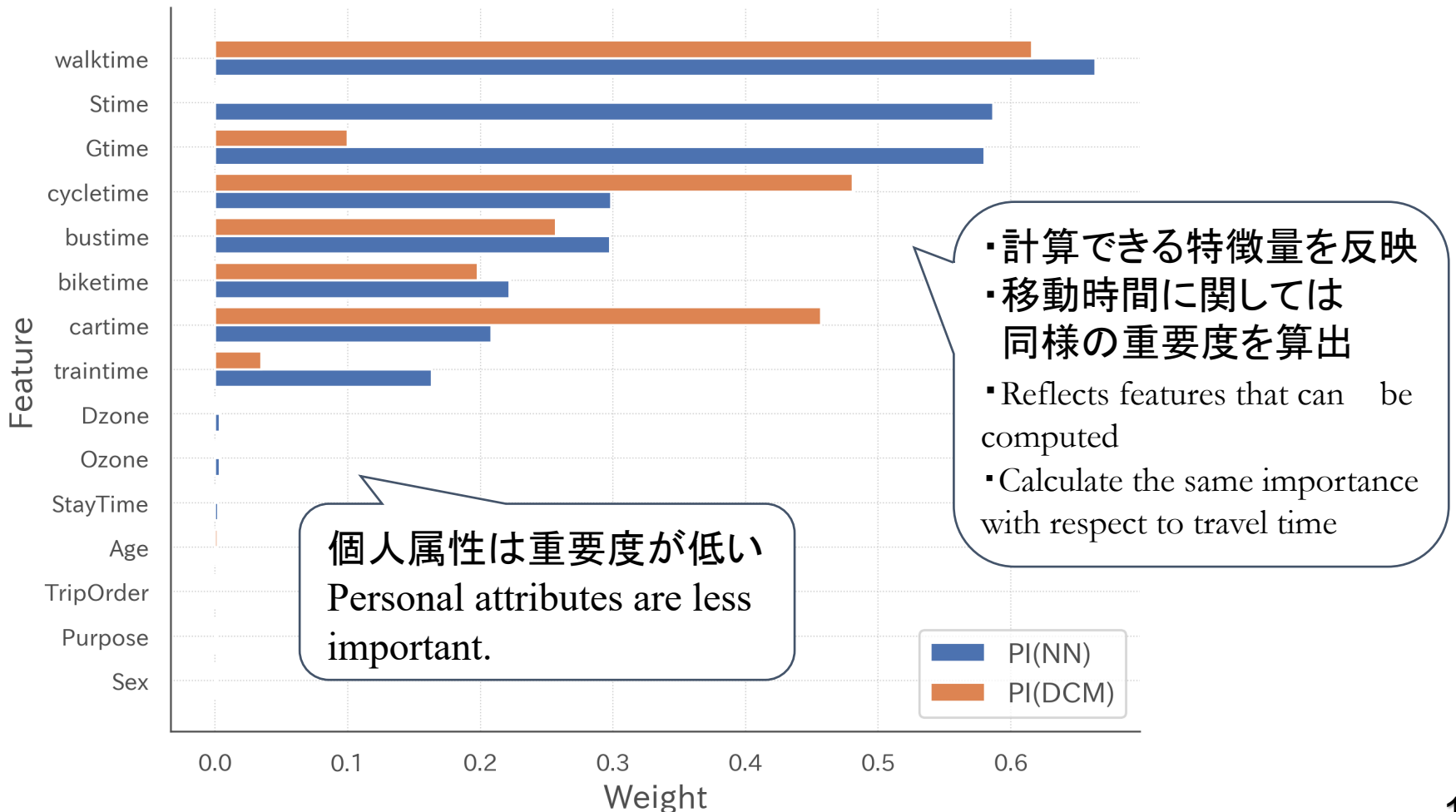
Permutation Importance(PI)

• 【DCM+NN】 と 【DCM】 で特徴量重要度を比較

Compare the PI of DCM+NN model and DCM

➤ DCMの重要度は、変数の有無による差を元に算出

Importance of DCM is calculated based on the difference with and without variables





離散選択モデルと機械学習の比較

DCM

NN

DCM+NN

解釈性
Explainability

- 変数パラメータ
Value of Parameters
- **弾力性**
Elasticity
- **感度分析**
Sensitivity analysis



- **特徴量重要度**
feature importance
- 特徴量の**感度分析(非線形)**
Sensitivity analysis of features(non-linear)
- **グループの異質性**
Group heterogeneity



- **特徴量重要度**
feature importance
- 特徴量の**感度分析(線形)**
Sensitivity analysis of features(linear)
- **グループの異質性**
Group heterogeneity



記述性能
Accuracy

- 処理速度が遅い
processing speed : slow
- 的中率 **0.624**
accuracy 0.624



- **高精度**
high precision
- 処理速度が速い
processing speed : fast
- 的中率 **0.752**
accuracy 0.752



- **高精度**
high precision
- 処理速度が速い
processing speed : fast
- 的中率 **0.557**
accuracy 0.557



理論性
theoreticality

- **効用理論に基づくルール**
Rules based on utility theory



- **損失関数の最小化**
Minimize loss function
- 再現性は高い
High reproducibility



- **損失関数の最小化**
Minimize loss function
- 再現性は高い
High reproducibility



- 離散選択モデルと機械学習の「**使い分け**」と「**併用**」
The "proper use" and "user" of DCM and machine learning

使い分け proper use

- 交通政策の決定など**理論性**が必要な場合は離散選択モデル
Use DCM when **theoreticality** is required, such as in transportation policy decisions
- ABMの計算など**記述性能**が求められる場面においては機械学習
Use machine learning when **descriptive performance** is required, such as ABM

併用 Using both

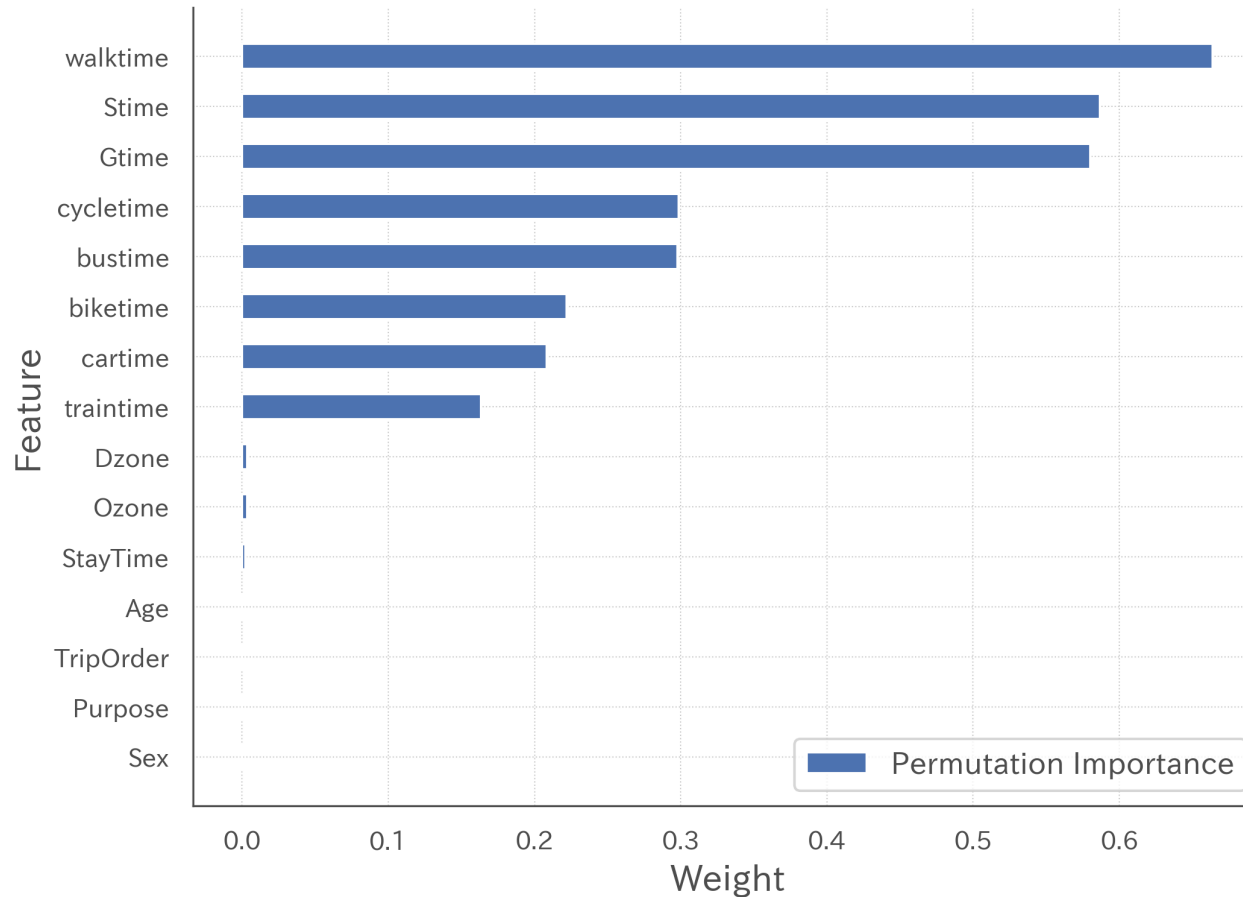
- 理論的（離散選択モデル）に導かれた結果と
精度が高いモデル（NN）から得られた結果の2つを用いる
- Using two results, one from the DCM and the other from the Neural Network
- **理論と精度**の両面から行動分析を行うことでより正確に評価できる
- More accurate policy evaluation is possible by analyzing behavior in terms of both **theory** and **precision**



予備スライド



- DCM+NNの交通手段選択モデルにPIを適用
 - 特徴量の重要度を計算



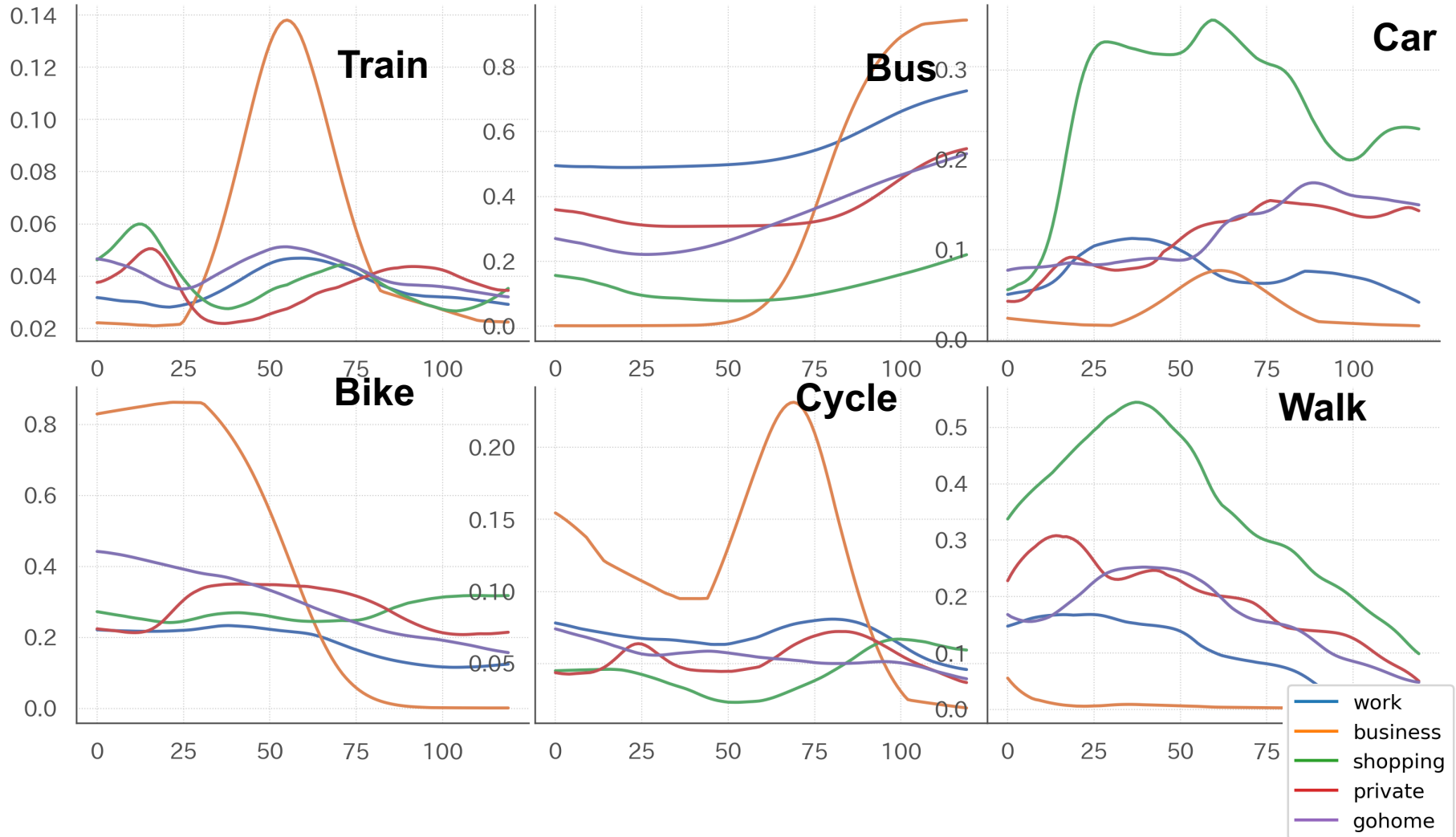
- どの特徴量を優先的に入れるかを判断できる
- 離散選択モデルでも変数の選択に用いることが可能



Conditional Partial Dependence(CPD)



- CPDを適用し、グループごとに特徴量と予測値の関係性を分析
 - 移動時間(横軸)と選択確率(縦軸)の関係 (非線形)



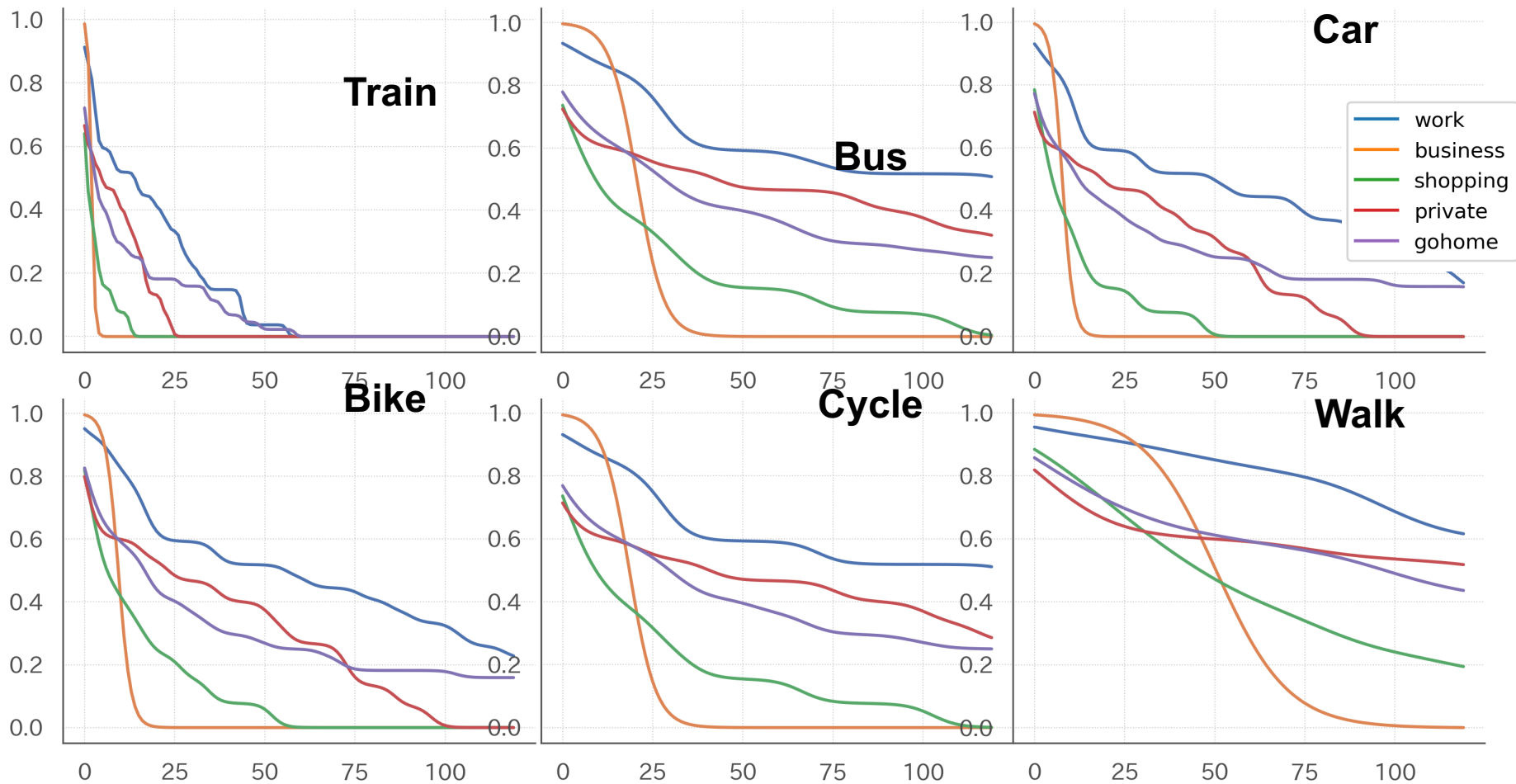


Conditional Partial Dependence(CPD)



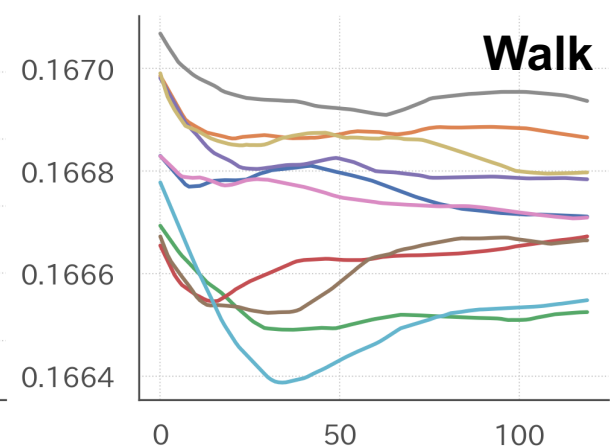
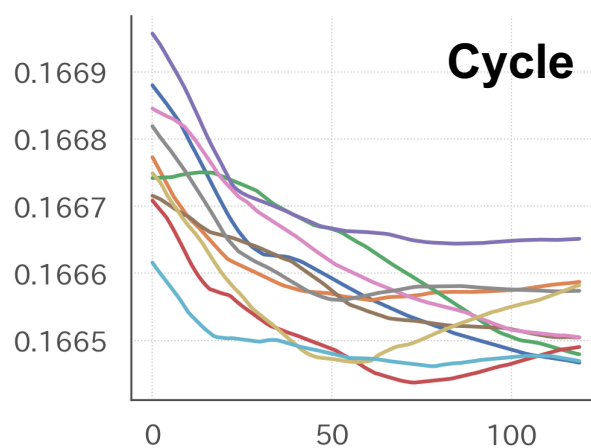
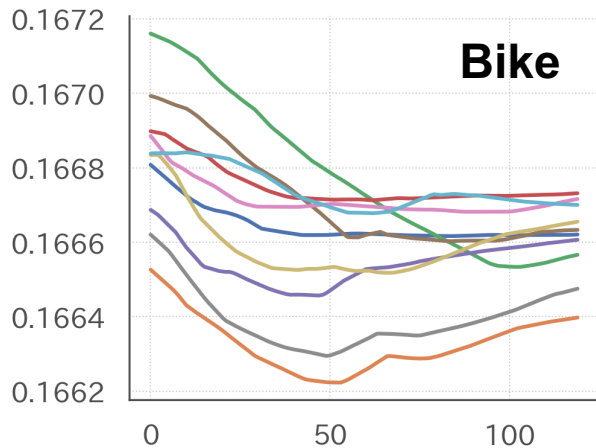
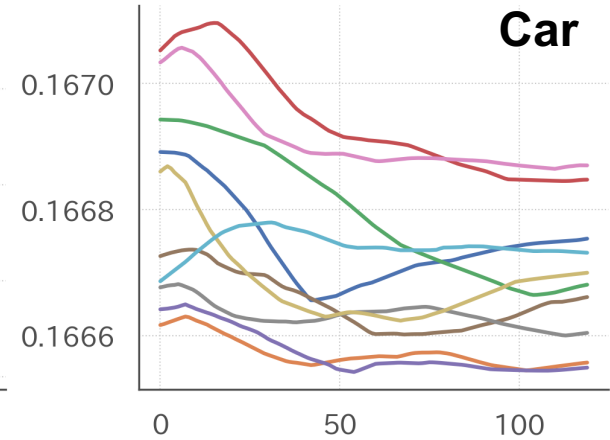
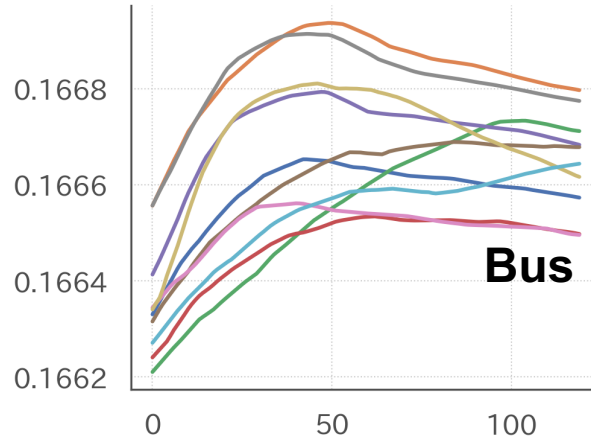
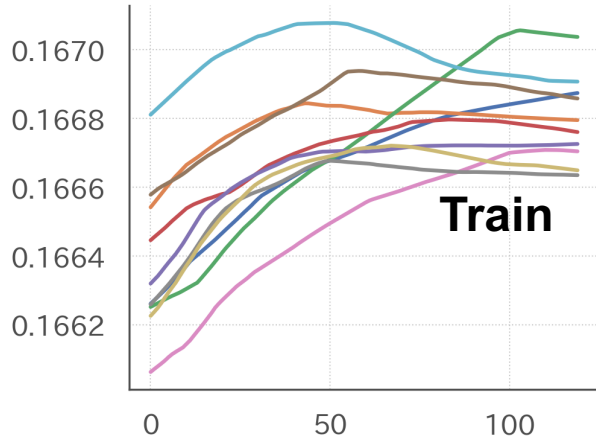
- CPDを適用し、グループごとに特徴量と予測値の関係性を分析

➤ 移動時間(横軸)と選択確率(縦軸)の関係 (線形)





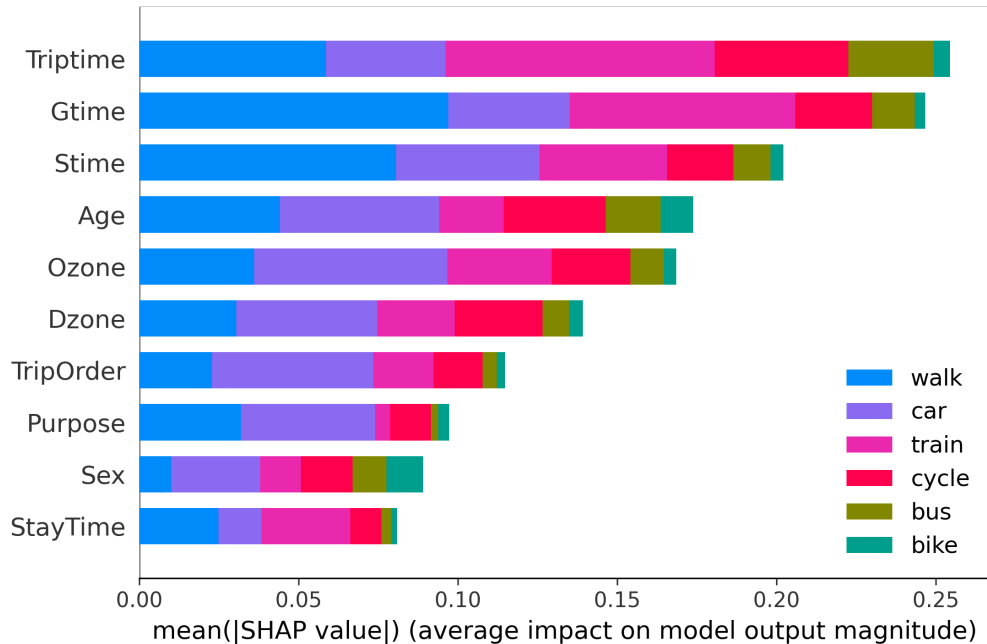
- 個人ごとの特徴量と予測値の関係性を分析
 - 移動時間(横軸)と選択確率(縦軸)の関係





• Summary_plot

➤ 目的変数に対する変数の寄与度を可視化したもの



PIと似ているが

「何を持って重要な特徴量とするか」
の定義の部分が異なっている

PI

モデルの性能低下に基づく重要度

SHAP

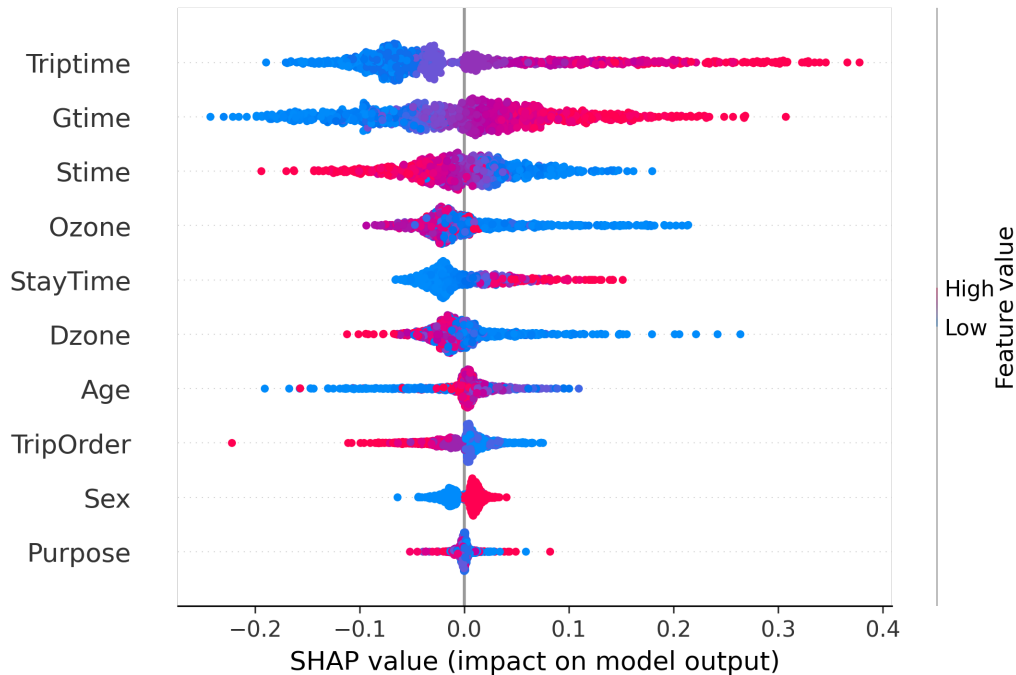
目的変数にどれだけ影響を与えるか



- 選択肢全体で見てもPIと同じような変数の重要度が大きい
- 選択肢によって重要度の順序は異なっている



- **Summary_plotに特徴量の値を反映させたもの**
 - 目的変数に対する変数の寄与度を可視化したもの



「鉄道」における特徴量重要度

- 特徴量の値が大きければ**赤く**
特徴量の値が小さければ**青く**なる
- 上位の特徴量ほど重要度は大きい

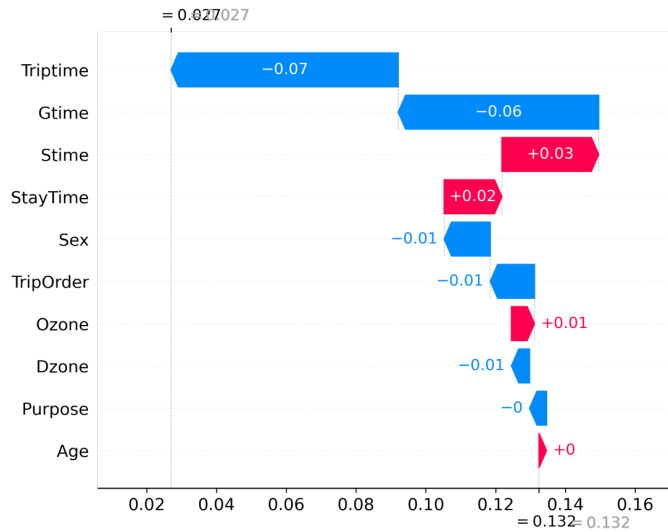
- 移動時間の場合
移動時間が大きければ
SHAP値も大きくなる



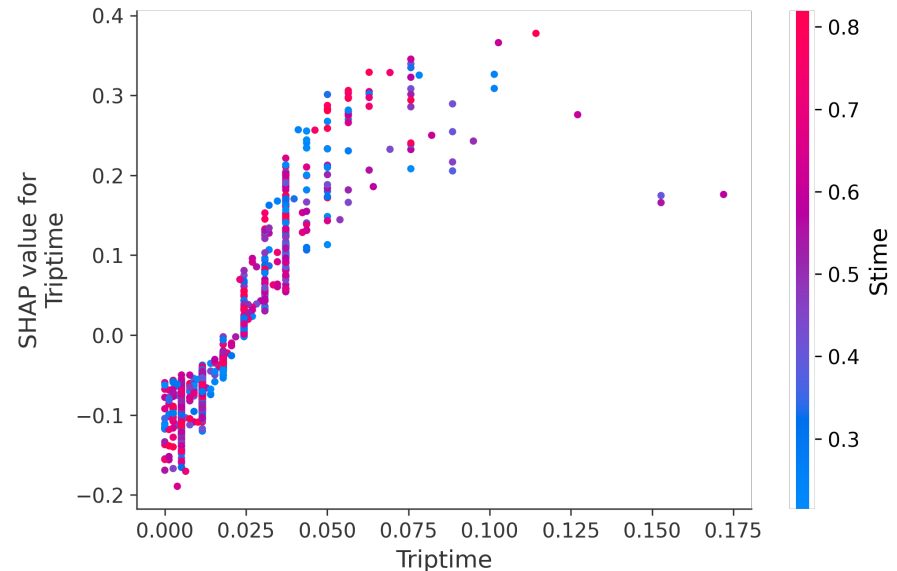
鉄道を選択する確率が大きくなる



• Waterfall_plotとDependence_plot



Waterfall_plot



Dependence_plot

【どのような経緯で選択確率を算出したか】

- ある個人の鉄道に対する選択確率推移
- 効用関数のパラメータと同様の解釈をすることが可能
- 離散選択モデルではデータ全体のパラメータであったが、SHAPでは個人のパラメータを見ることが可能

【特微量とSHAP値の関係性】

- 鉄道における移動時間とSHAP値の関係
- カラーバーにさらに特微量を指定可能
- ある値までは線形的だがそれ以降はばらつきが見られる
- 車の費用が大きく移動時間が大きいと鉄道を選択しやすくなる



マクロな
解釈手法

PI

SHAP

PD

SHAP

ミクロな
解釈手法

CPD

SHAP

SHAPに関しては

- インスタンスごとの**ミクロな手法**
- 適度な粒度で集計・可視化を行うことでPI・PDのように**マクロな手法**として利用可能

「弱い」使い方
比較的安全

モデルのデバック

- 事前知識と整合的か、想定外の挙動はないか

モデルの振る舞いを解釈

- モデルは特徴量Aを重視している
 - 特徴量が大きくなると予測値が大きくなる
- あくまでモデルの一側面であり間違った解釈をする可能性

因果関係の探索

- モデルの振る舞いを因果関係として解釈
- 厳密な因果推論の手法を合わせて使用するべき

「強い」使い方
注意が必要



Permutation Importance による特徴量の重要度

- モデルにおける特徴量重要度を計算する指標
- ある特徴量をシャッフルし予測精度が悪化する → 重要な特徴量

ランダムにシャッフル

Person	年齢	性別	移動時間		Person	年齢	性別	移動時間
1	24	M	15	➤	1	24	M	30
2	15	F	30		2	15	F	20
3	30	M	35		3	30	M	15
4	46	F	20		4	46	F	35

- 各変数に対して同様の処理を行うことで計算可能
- ランダム化された特徴量で予測したモデルの誤差が大きくなればその変数はモデルにおいて重要である
 - 損失関数の差分が特徴量重要度となる
- モデルの精度を高めるための重要度であり因果的な解釈は不適切
 - 重要度が高い変数に対してより深い分析を行うことが重要



それぞれの特徴量がモデルの予測値にどのように影響を与えるのか

- 特徴量と予測値の平均的な関係を計算する
- 特徴量が大きくなったときモデルの予測値は大きくなるのか
その関係は非線形なのかについて判断が可能
- 他の特徴量を固定しある特徴量だけを動かし感度を分析する

X_0	X_1	X_2
1	2	5
2	7	2
3	3	4



X_0	X_1	X_2	予測
1	2	5	$f(1,2,5)$
1	7	2	$f(1,7,2)$
1	3	4	$f(1,3,4)$



$$\frac{1}{3} \{f(1,2,5) + f(1,7,2) + f(1,3,4)\}$$

X_0	X_1	X_2	予測
2	2	5	$f(2,2,5)$
2	7	2	$f(2,7,2)$
2	3	4	$f(2,3,4)$



$$\frac{1}{3} \{f(2,2,5) + f(2,7,2) + f(2,3,4)\}$$

X_0	X_1	X_2	予測
3	2	5	$f(3,2,5)$
3	7	2	$f(3,7,2)$
3	3	4	$f(3,3,4)$



$$\frac{1}{3} \{f(3,2,5) + f(3,7,2) + f(3,3,4)\}$$

予測結果を平均



- 協力ゲーム理論における**Shapley Value**を機械学習に反映
 - Shapley Value**: 利益配分の基準として用いられる指標
 - 複数のプレイヤーが協力して得た利益を各プレイヤーの貢献度に応じて配分する際に用いられる

参加者	報酬
A	6
B	4
C	2
A,B	20
A,C	15
B,C	10
A,B,C	24

協力ゲーム理論



機械学習モデルの
各特徴量に置き換え









特徴量	予測値変化
X_0	1.0
X_1	0.7
X_2	0.3
X_0, X_1	2.5
X_0, X_2	1.5
X_2, X_3	1.0
X_0, X_1, X_2	3.0

機械学習

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S))$$

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} (f_x(S \cup \{i\}) - f_x(S))$$



	離散選択モデル	機械学習
記述性能	 <ul style="list-style-type: none">・ 選択肢が多くなると悪化・ 処理速度が遅い	 <ul style="list-style-type: none">・ 精度が高い・ 計算速度が速い
解釈性	 <ul style="list-style-type: none">・ 全体のパラメータしか得られない	 <ul style="list-style-type: none">・ マクロでもミクロでも可・ 個人ごとのパラメータがわかる
データセット	 <ul style="list-style-type: none">・ 手間がかかる・ 実選択肢以外も考える必要性	 <ul style="list-style-type: none">・ 実選択肢のみのデータセットで十分
理論性	 <ul style="list-style-type: none">・ 効用による合理的なルール	 <ul style="list-style-type: none">・ 損失関数の最小化・ 精度が高いため再現性は高い



- 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手法
 - マクロとミクロの両方の視点からの解釈が可能



SHAP

