19 Sep. 2023 @ UT

Machine Learning for Behavior Model

How to collaborate DCM and ML ?

Tokyo University of Science

Hideki YAGINUMA

yaginuma@rs.tus.ac.jp

DCM vs ML

To develop **more better** infrastructure and services, we are conducting analysis based on behavior models.

- **Prediction Accuracy**: issue of model parameter estimation
- **Model Interpretability**: issue of understanding model (behavior)

Discrete choice model

$$P(i) = \frac{\exp(\mu V_i)}{\sum_{j \in C} \exp(\mu V_j)}$$

Source: Small data
Function: Liner
Accuracy(Model fitting): Mid（Low?）
Interpretability : High
Good understanding travel behavior ?

Machine Learning

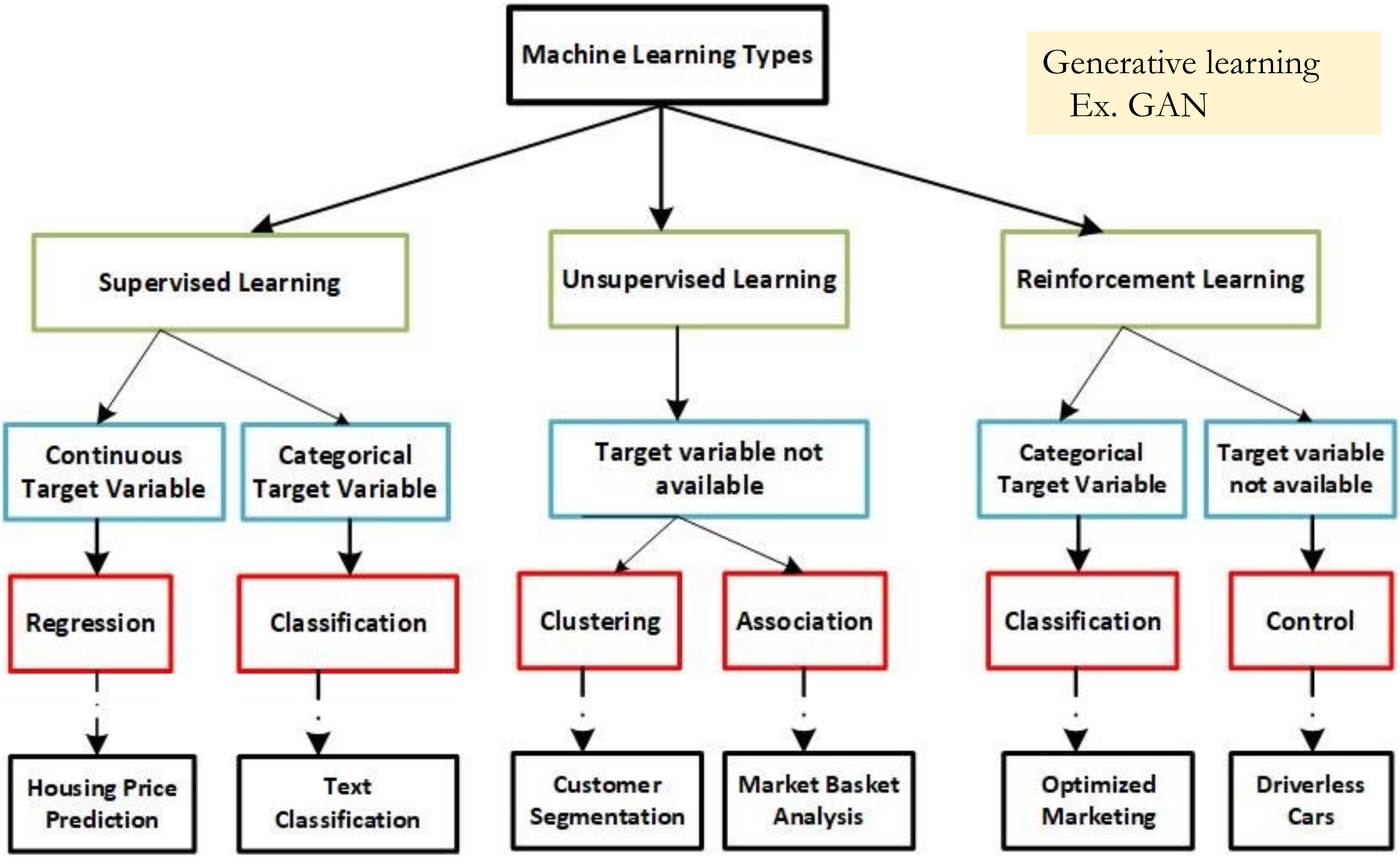
Black Box

Source: Big data
Function: Non-Liner
Accuracy(Model fitting): High
Interpretability: Low
NOT understanding travel behavior ?

Considering the possibility of collaboration between DCM and ML from the perspective of Aaccuracy and Interpretability

Overview of ML

Generative learning
Ex. GAN



https://ieeexplore.ieee.org/document/8612259

How to collaborate DCM and ML ?

Application to research field

- You can (not) apply behavior model-

Our on-going research topics

1. AI based road traffic observation system
   - Automatic traffic data collection using CCTV camera for whole national roads managed by the MLIT.
   - Apply to CNN（YOLO）and transfer-learning by original data.
   - Development of annotation-free self-learning algorithm.

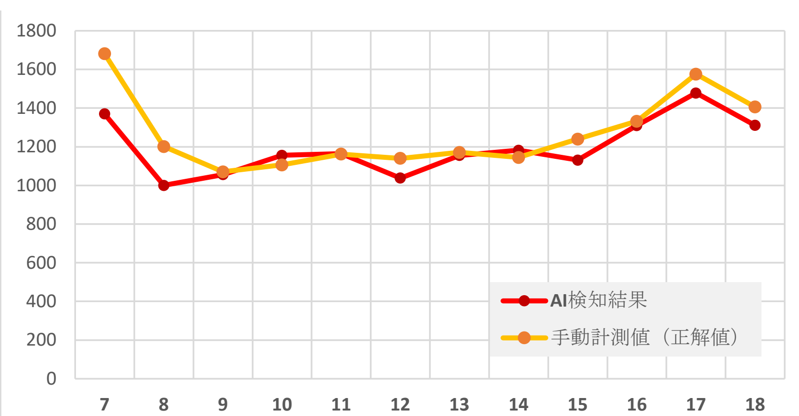
2. Data generation by multi-data fusion
   - Fusion of multiple data source such as national census, traffic count and probe.
   - OD data generation by Gaussian Process model.
   - Activity data generation by GAN.

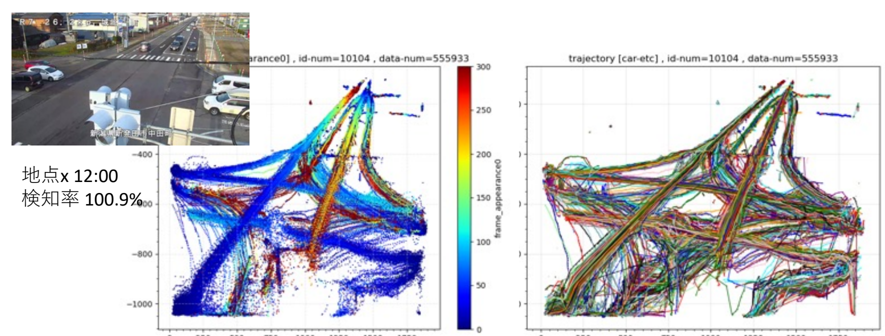
3. NN based choice model and Activity simulation
   - Non-liner choice model by neural network (NN)
   - Consider explainability of Machine learning.
   - Develop the NN type activity simulator and application to Tokyo Metropolitan area.

Summary of Topic 1

1. AI based road traffic observation system

- Apply to CNN（YOLO + DeepSORT）and transfer-learning by original data.
- Obtained high accuracy of over 95%

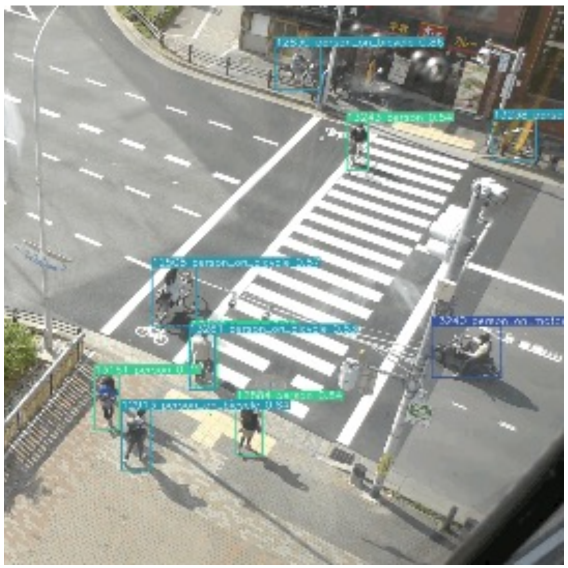


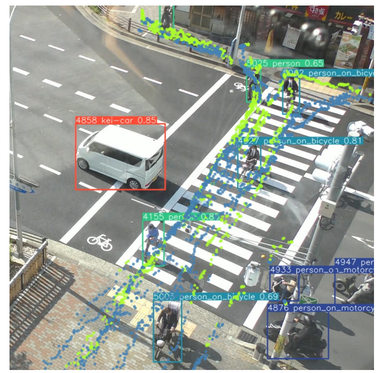


Summary of Topic 1

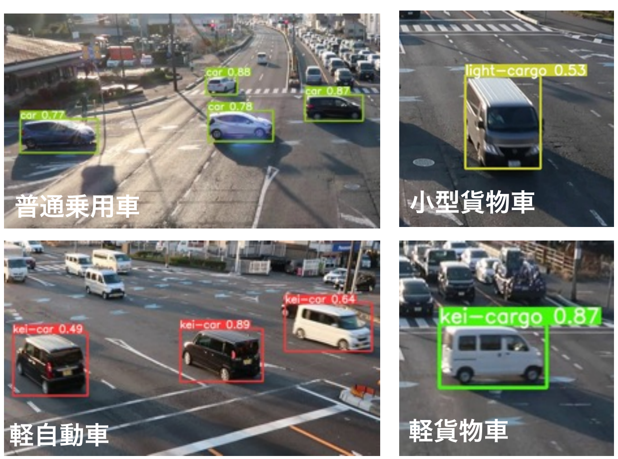
1. AI based road traffic observation system

- High accuracy detection of car type, bicycle, pedestrian
- Optimize count line in CCTV



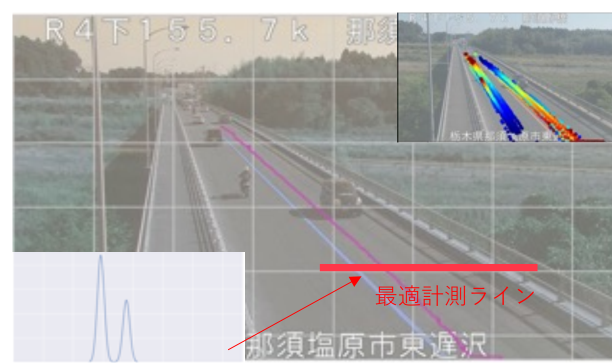


$$s\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$



車種区分（7種類）

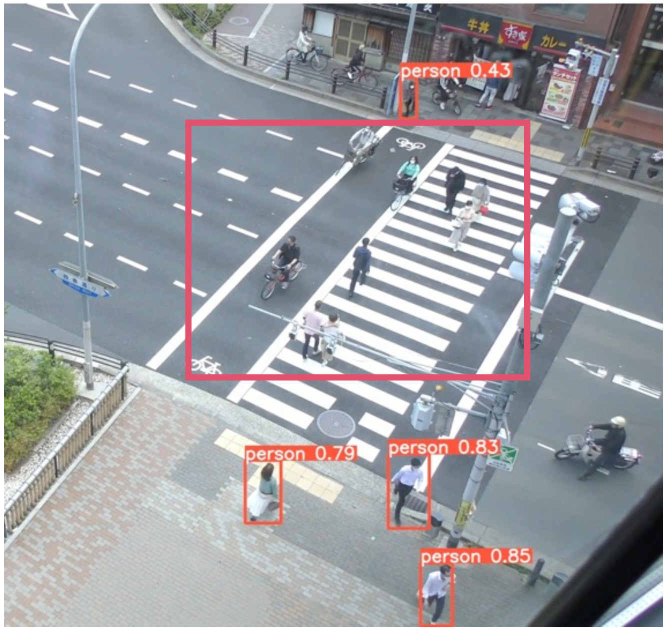
1. 普通乗用車
2. 軽乗用車
3. 軽貨物車
4. 小型貨物車
5. 普通貨物車
6. バス
7. 特殊車両



Summary of Topic 1

1. AI based road traffic observation system

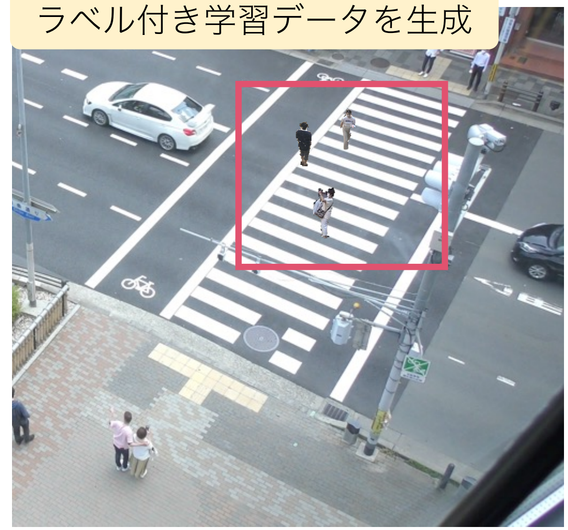
- Development of annotation-free self-learning algorithm.



横断歩道上で歩行者
の検知精度が低い

転移学習

ラベル付き学習データを生成



背景画像に合成

AIモデルで学習ターゲット
のラベルを持つ領域を抽出

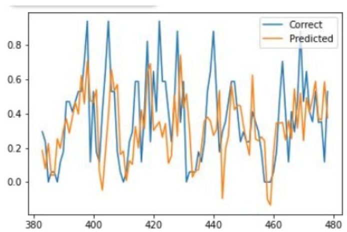
Mask R-CNNを使用

ラベルと輪郭情報を取得

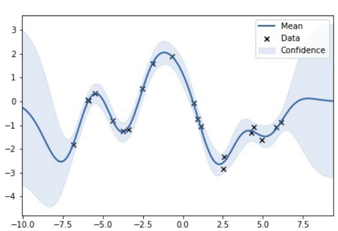
2. Data generation by multi-data fusion

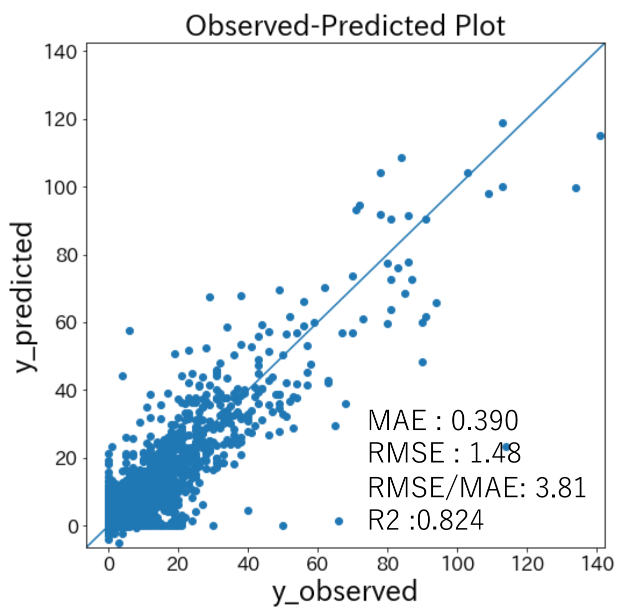
- Fusion of multiple data source such as national census, traffic count and probe.
- Link volume estimation by GCN+LSTM.
- Realtime OD data generation by Gaussian Process Regression.
- Activity data generation by GAN.

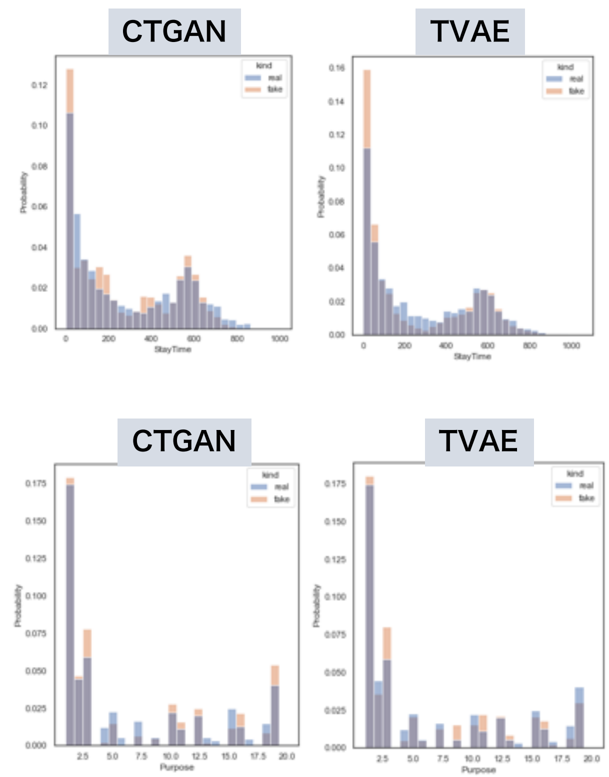
GCN+LSTM
Generate link Vol



GPR
Generate OD Vol







Our on-going research topics

1. AI based road traffic observation system
    - Automatic traffic data collection using CCTV camera for whole national roads managed by the MLIT.
    - Apply to CNN（YOLO）and transfer-learning by original data.
    - Development of annotation-free self-learning algorithm.

2. Data generation by multi-data fusion
    - Fusion of multiple data source such as national census, traffic count and probe.
    - OD data generation by Gaussian Process model.
    - Activity data generation by GAN.

3. **NN based choice model and Activity simulation**
    - **Non-liner choice model by neural network (NN)**
    - **Consider explainability of Machine learning.**
    - **Develop the NN type activity simulator and application to Tokyo Metropolitan area.**

Main Contents

1. Background and Objective of the Study
2. What is XAI(explainable AI)?
3. Application of XAI
4. Change Neural Network structure
   and comparison of MNL and NN
5. Activity Simulation in Tokyo area

Contents

1. Background and Objective of the Study

2. What is XAI(explainable AI)?

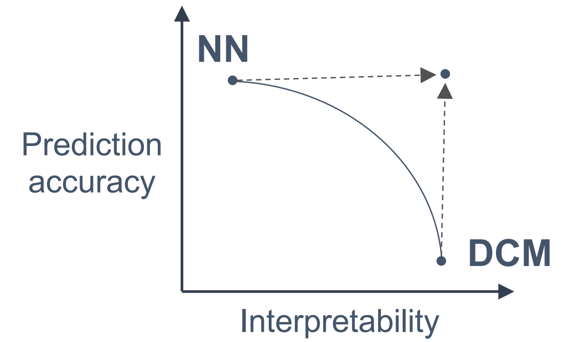
3. Application of XAI

4. Change Neural Network structure and comparison of MNL and NN

5. Activity Simulation in Tokyo area

Background & Objective

Travel Behavior model



Discrete Choice Model (DCM)

- Highly interpretable
- It has been used from the viewpoints of "interpretability" and "logicality"

Machine Learning (ML)

- High descriptive performance
- Black box models, lack of interpretability
- Development of interpretive indicators(Explainable AI: XAI) of sensitivity and predictive basis
  - PI, PD, CPD, SHAP

Applicability of ML in Travel Behavior Models

- Development of travel behavior model with neural network
- Applying XAI to understanding travel behavior

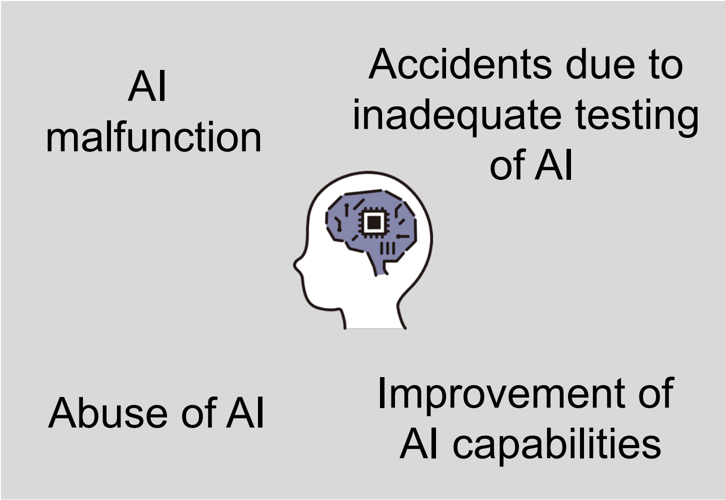
Contents

1. Background and Objective of the Study
2. **What is XAI(explainable AI)?**
3. Application of XAI
4. Change Neural Network structure and comparison of MNL and NN
5. Activity Simulation in Tokyo area

Explainable AI(XAI)

Explainable AI(XAI) is a tool for interpreting machine learning

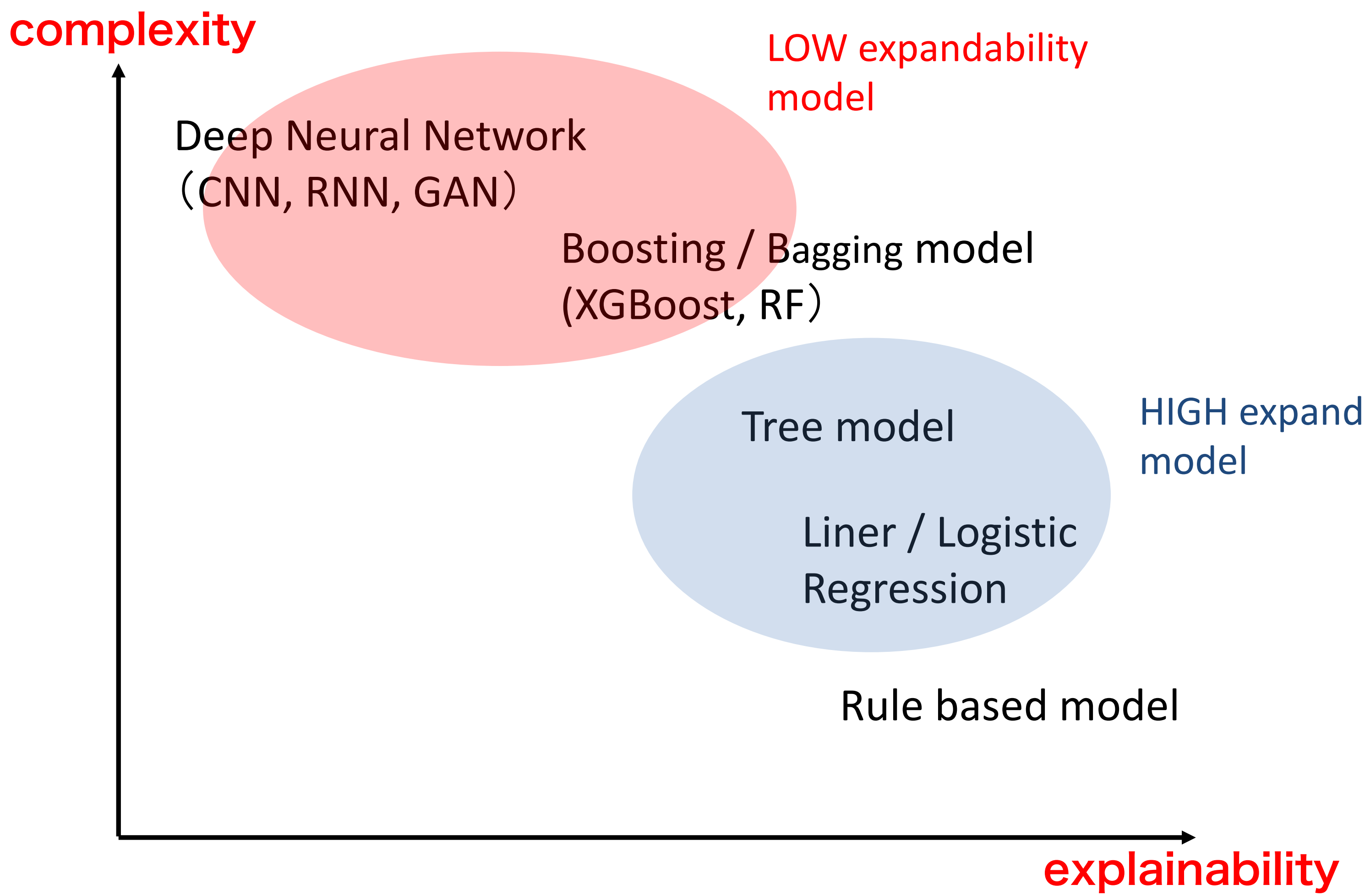
- Mainly focused on improving accuracy in “regression” and “classification”
- It can add interpretability to black box models



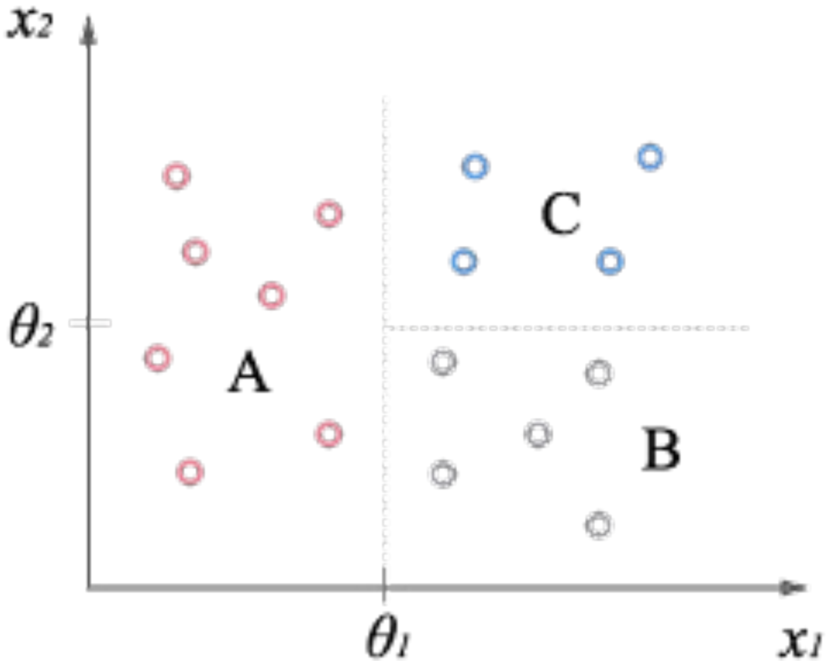
- In meeting fairness, transparency, and accountability,
  - Necessity to know about prediction basis and learning process of the model
    - The basis for the output of the model
    - How much the feature values affect the prediction values
- Developing a mode choice model with ML, and applying XAI to it.
- Comparing **interpretability** and **accuracy** with discrete choice model

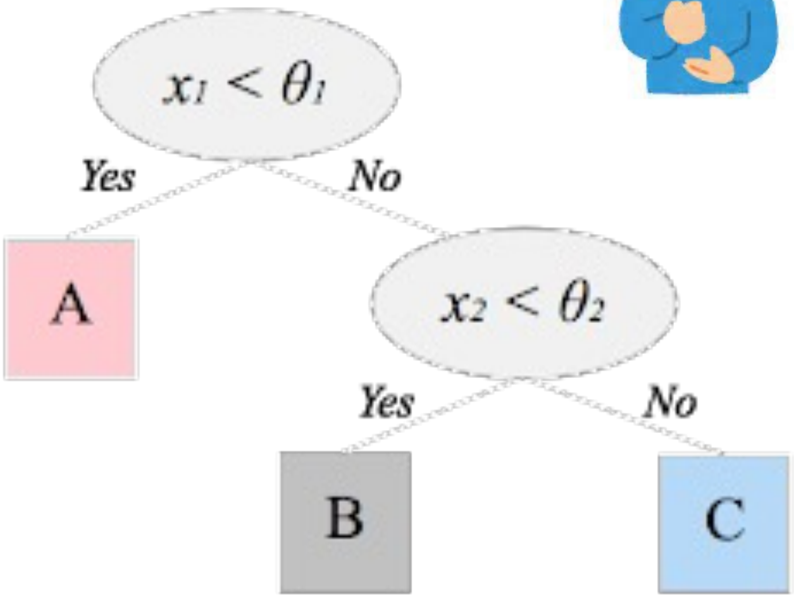
Explainability of ML

In machine learning, there is a **tradeoff** between the **complexity** (= accuracy) of the problem and its **explainability**.

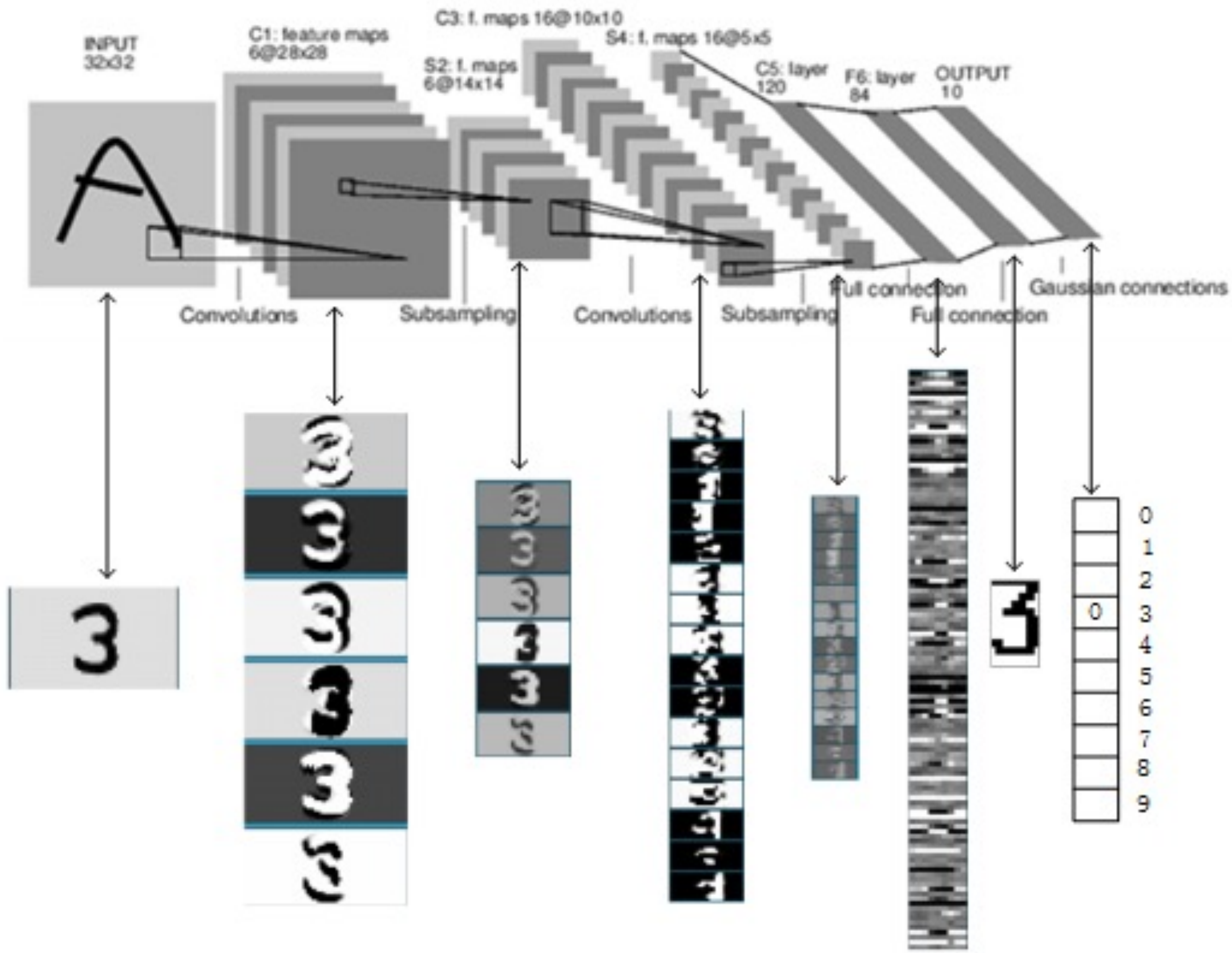


Tree model



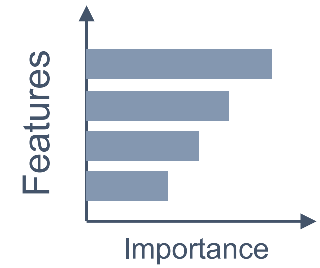


CNN model


INPUT 32x32
C1: feature maps 6@28x28
S2: f. maps 6@14x14
C3: f. maps 16@10x10
S4: f. maps 16@5x5
C5: layer 120
F6: layer 84
OUTPUT 10
Convolutions
Subsampling
Convolutions
Subsampling
Full connection
Full connection
Gaussian connections
0
1
2
3
4
5
6
7
8
9

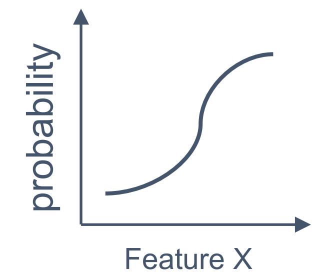
Explainable AI(XAI)

(1)Permutation Importance(PI)



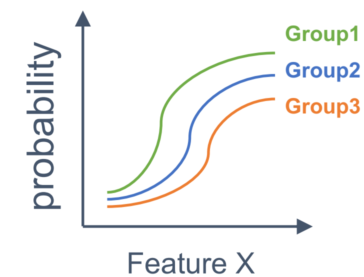
- Calculate **the feature importance** of the model by randomly shuffling features.
- Determine which features are important for the accuracy of the model.

(2)Partial Dependence(PD)



- Calculate **the average relationship between features and predictions** to see how a particular feature affects the model's predictions.
- Determine whether the feature and the predicted value are proportional or inversely proportional, linear or nonlinear.

(3)Conditional Partial Dependence(CPD)



- Indicators grouping PD.
- determine heterogeneity by group.

(4)SHapley Additive exPlanations(SHAP)

- Calculate **the contribution of the features to the predictions**.
- Shapley Value in Cooperative Game Theory is Reflected in Machine Learning.
- Shapley Value: An index used as a basis for profit sharing.
- interpret why model outputs such predictions.

Comparison of Explainable AI (XAI)

Comparison of analytical granularity

Macro	PI	SHAP
	PD	SHAP
Micro	CPD	SHAP

About SHAP
- Support for micro and macro analysis
- Instance-by-instance micro analysis
- Can be used as a macro method like PI,PD by aggregating and visualizing
- Explain specific examples later

Comparison of usage

Weak usage
relatively safe

Model debugging
- Determine if it is consistent with prior knowledge or if there is unexpected behavior

Interpret black box models
- The model emphasizes feature A
- As the feature value increases, the predicted value increases

→One aspect of the model can be misinterpreted

Exploring causality
- Interpret model behavior as causality

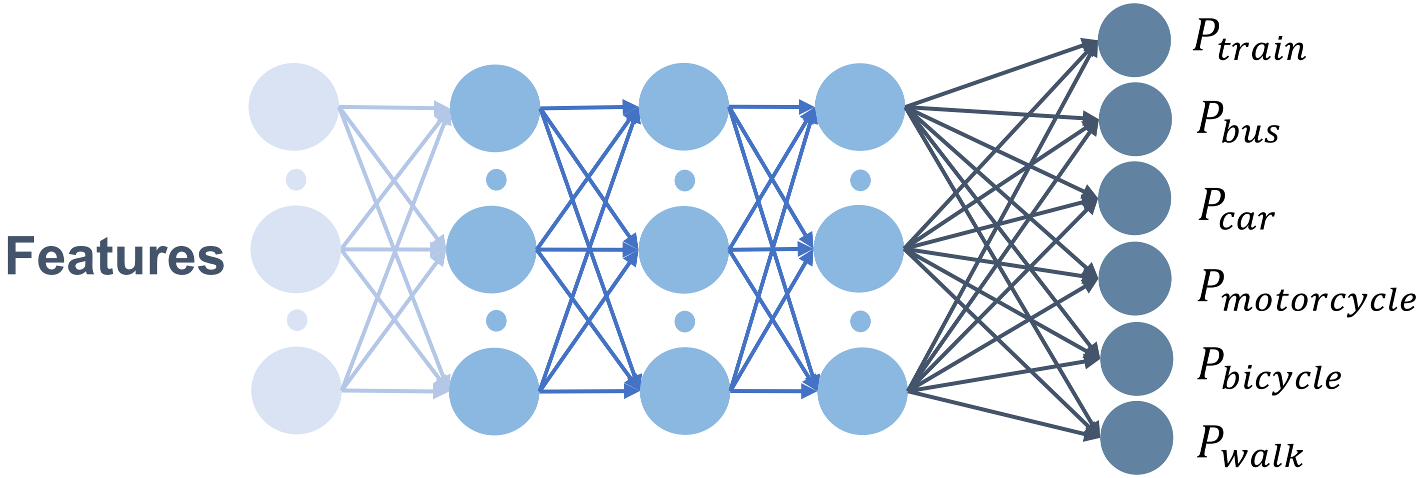
→Need to use methods of rigorous causal inference together

Strong usage
Caution is needed

Contents

1. Background and Objective of the Study
2. What is XAI(explainable AI)?
3. **Application of XAI**
4. Change Neural Network structure and comparison of MNL and NN
5. Activity Simulation in Tokyo area

Mode choice model with Neural Network



Accuracy
0.945

- The middle layer has 4 layers.
- Number of units is in order of 100,100,50,10.

Input Layer **Tokyo Person Trip data**

Gender, Age, Occupation, Income,
Travel time, Stay time, Purpose,
Departure facility, Destination facility,
Departure zone, Destination zone,
Departure time, Destination time,
Number of trip, Each mode fare,
Each travel time

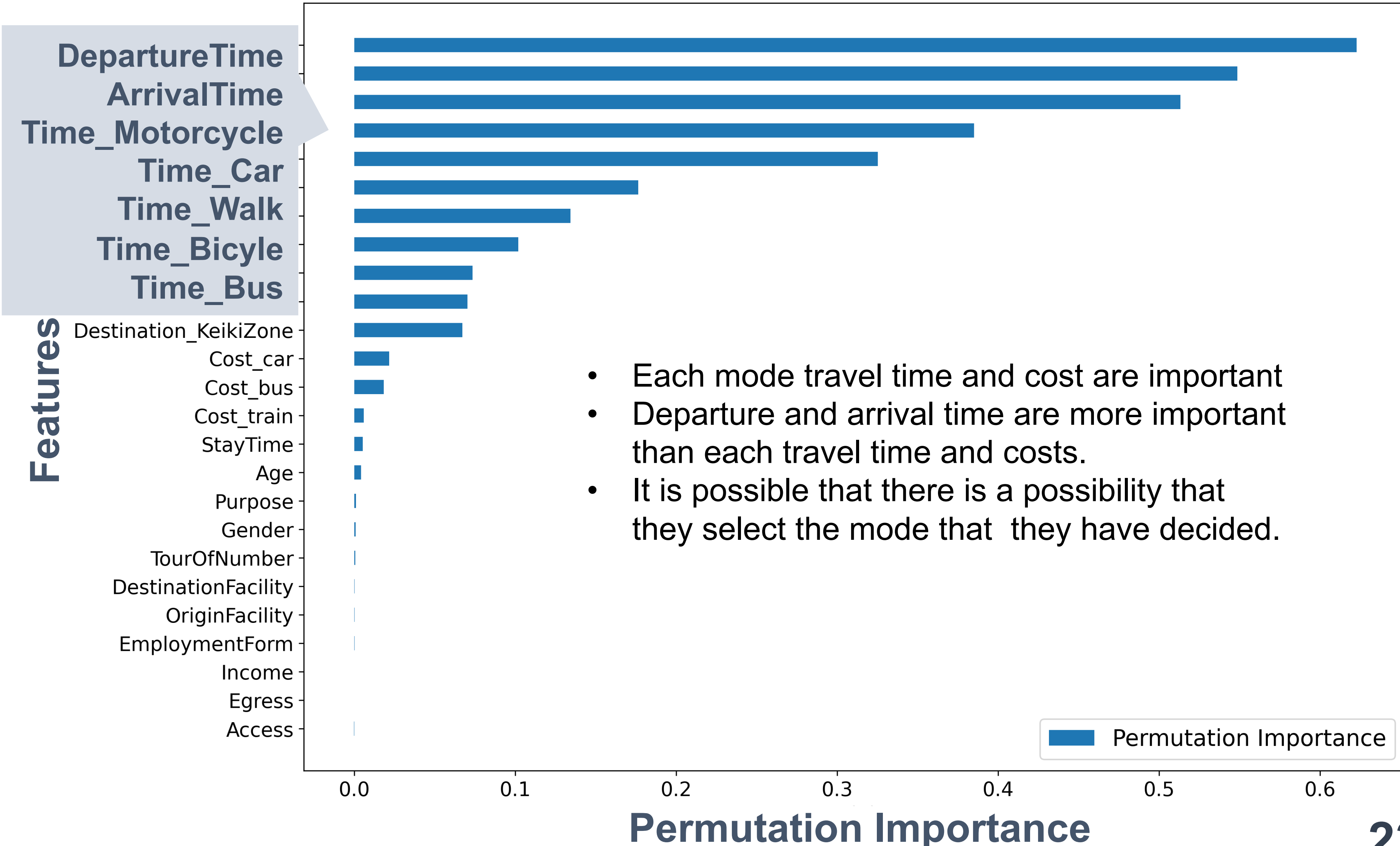
Output Layer

Choice probability of each mode
Train, Bus, Car,
Motorcycle, Bicycle, Walk

However, this model is **poor interpretability** in this states.
→ apply **XAI** to NN to give interpretability

Permutation Importance(PI)

- Applying PI to NN transportation mode choice model
  - ➢PI: calculate feature importance

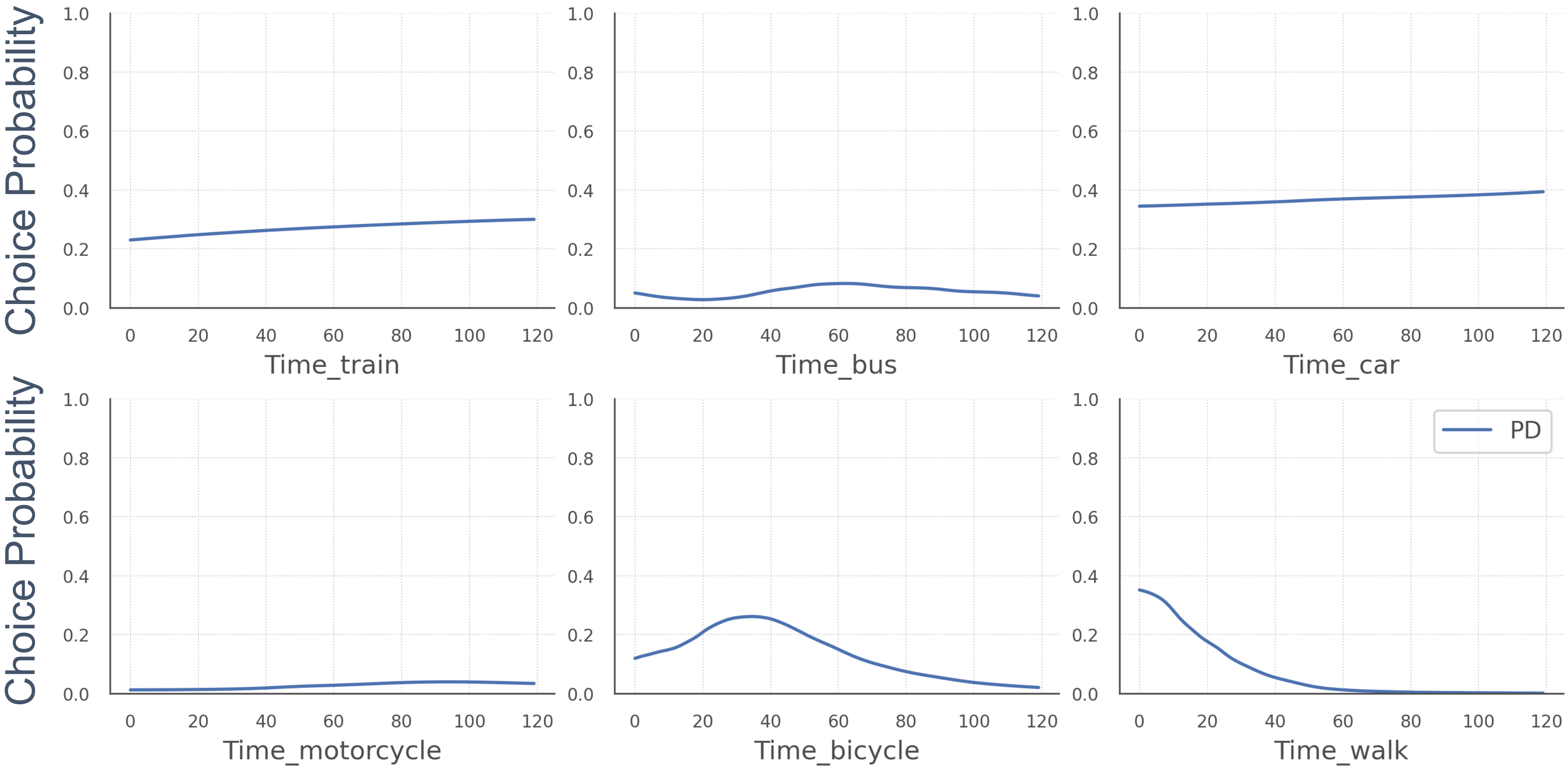


- Each mode travel time and cost are important
- Departure and arrival time are more important than each travel time and costs.
- It is possible that there is a possibility that they select the mode that they have decided.

Partial Dependence(PD)

PD: The relationship between features and predictions

➢horizontal axis: **Travel time(minutes)**, vertical axis: **Choice Probability**

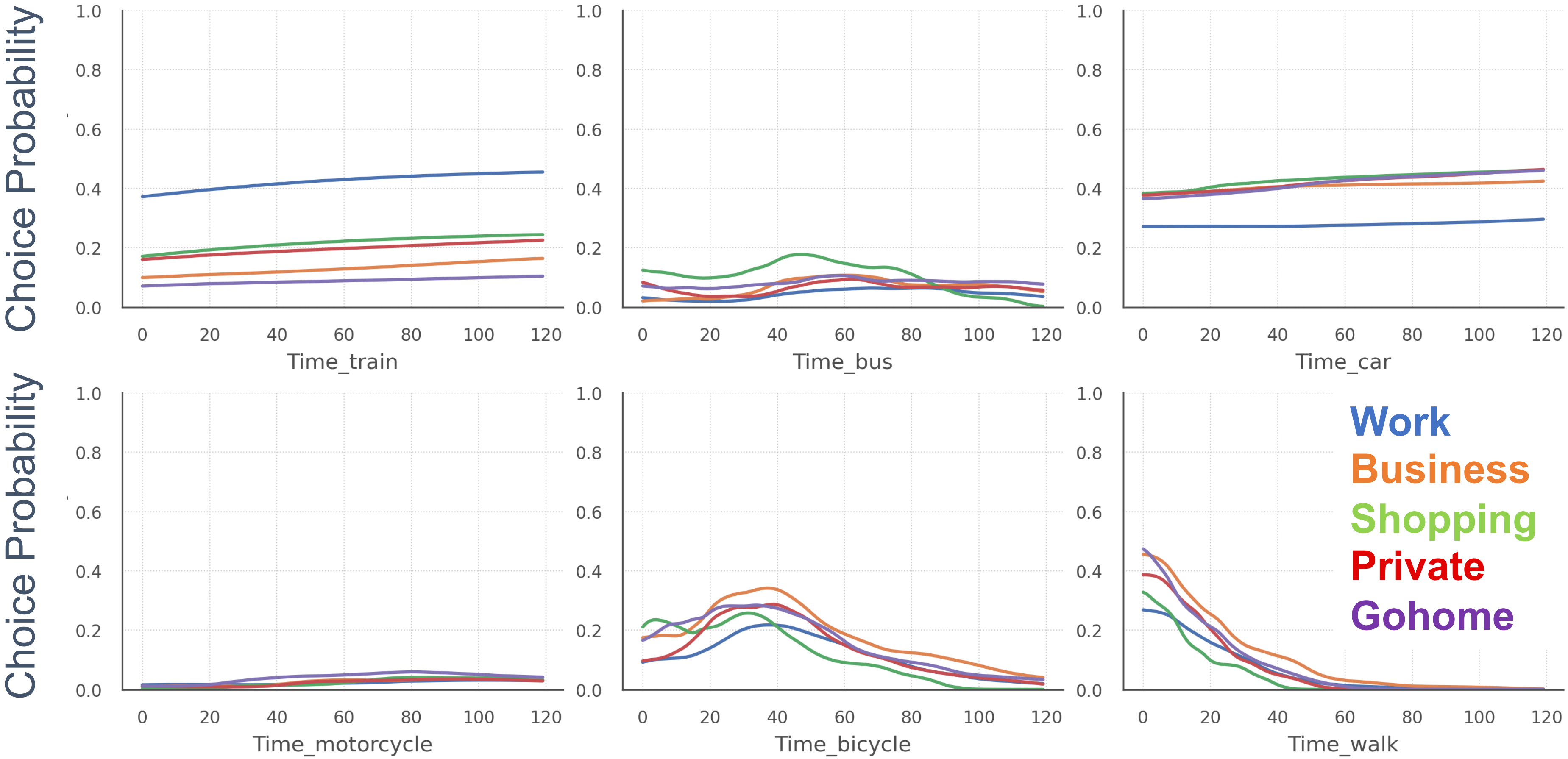


- Changing travel time of each mode.
- It was confirmed that there was a probability transition with an inflection point
- Capture behavioral changes that have an inflection point.

Conditional Partial Dependence(CPD)

Capturing the heterogeneity of each group

➢horizontal axis: **Travel time(minutes)**, vertical axis: **Choice Probability**

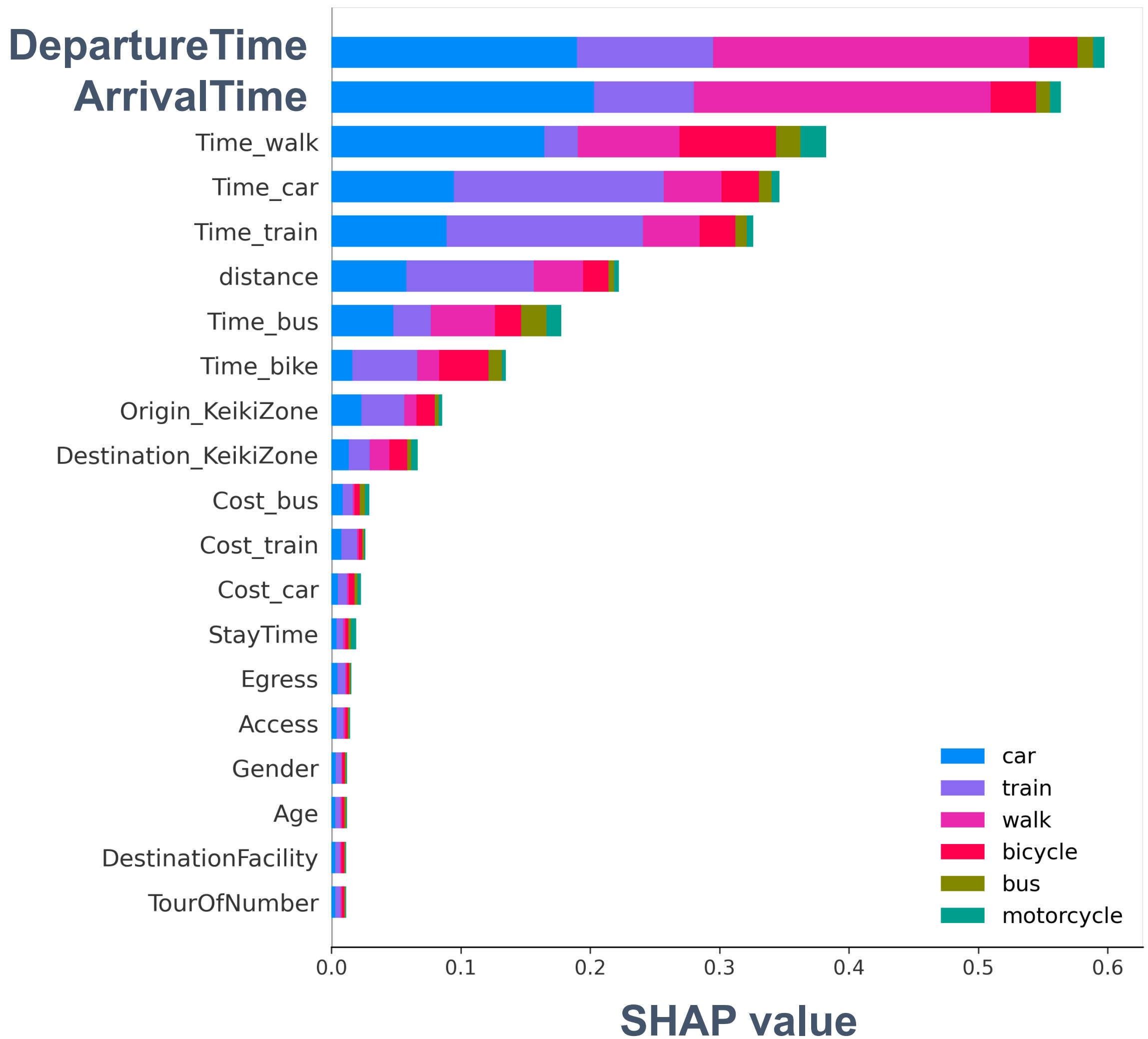


- Classifying trip purpose into 5.
- For example, people who commute to work(commuting to work or school) tended to choose trains more easily than other trip purposes.
- No significant heterogeneity was confirmed from the results

Shapley Additive exPlanations(SHAP)

Summary plot

➢Visualization of the contribution of features to the objective variable



Similar to PI, but different in the definition of what constitutes an important feature

PI

Importance based on model performance degradation

SHAP

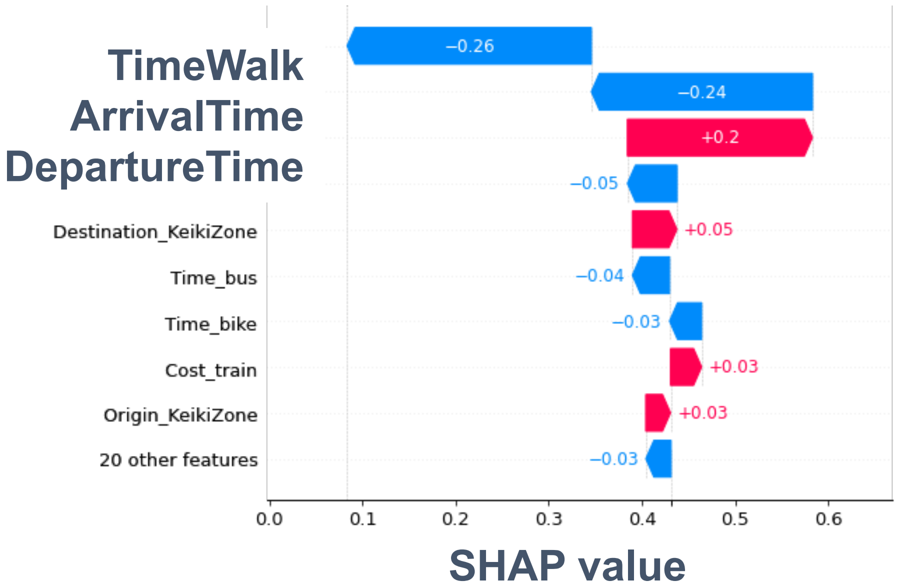
How much does it affect the objective variable

Trip departure and arrival times have a large effect, and it is thought that the fixed mode is selected.

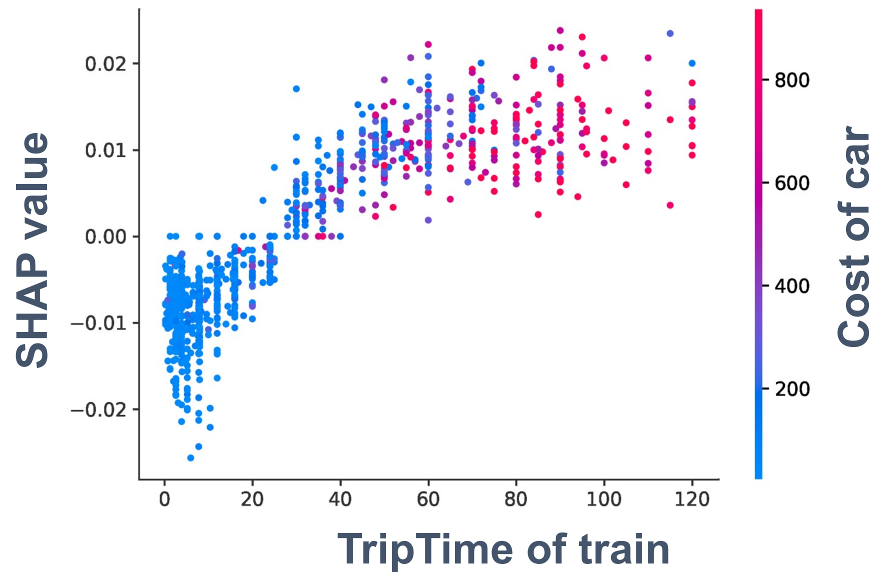
Shapley Additive exPlanations(SHAP)

“Waterfall plot” and “Dependence plot”

Waterfall plot



Dependence plot



Transition of an individual’s choices

- Probability transition of an individual’s choice of railroad
- In the discrete choice model, it was the parameter of the entire data, but in SHAP, it is possible to see the individual params.

Relationship SHAP and feature

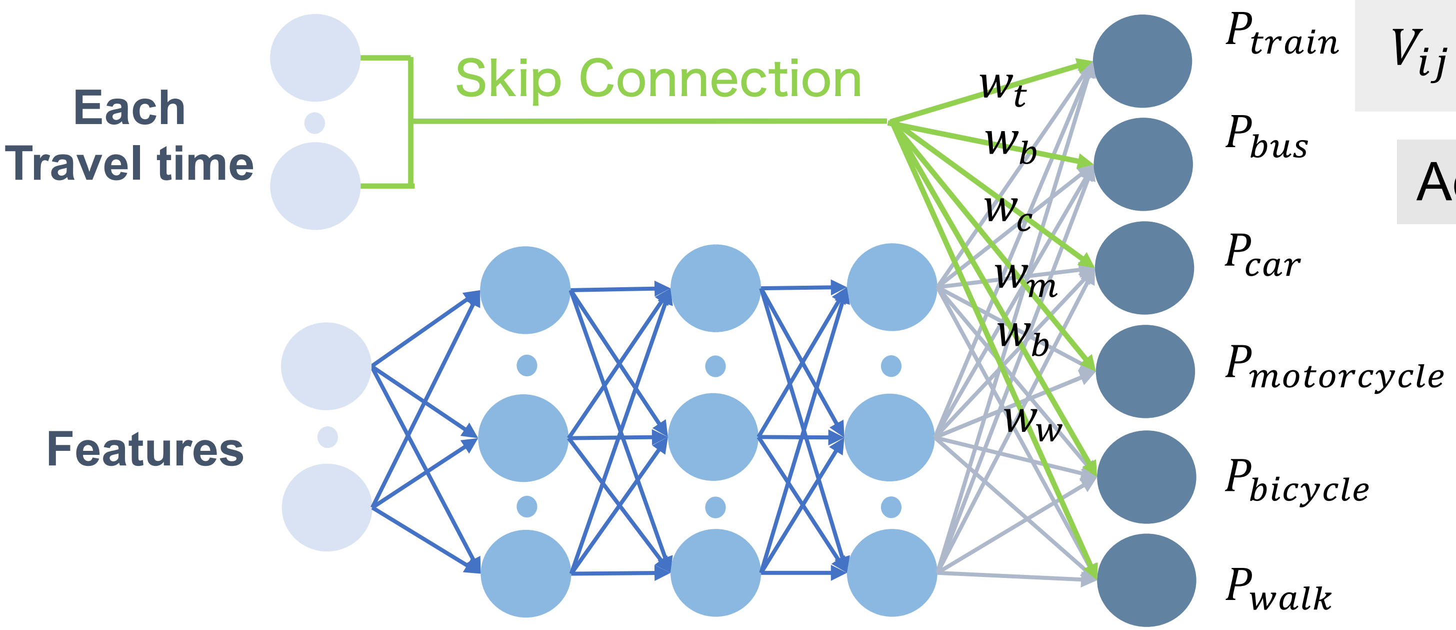
- The travel time of the train positively increases the choice probability.
- By adding another feature, we can analyze interrelationship
- It is easy to select the train for trips that the travel time of the train and cost of the car are large.

Contents

1. Background and Objective of the Study
2. What is XAI(explainable AI)?
3. Application of XAI
4. **Change Neural Network structure and comparison of MNL and NN**
5. Activity Simulation in Tokyo area

Networks reflecting elasticity

- Considering **elasticity** by changing the model structure like MNL
  - Sifringer, Brian, Virginie Lurkin, and Alexandre Alahi. "Enhancing discrete choice models with representation learning." Transportation Research Part B: Methodological 140 (2020): 236-261.



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

Accuracy 0.854

$$\begin{aligned} w_{train} &= 0.876 \\ w_{bus} &= 0.318 \\ w_{car} &= 0.635 \\ w_{mcycle} &= 0.277 \\ w_{bicycle} &= -0.104 \\ w_{walk} &= -0.927 \end{aligned}$$

- Avoiding travel time from the hidden layer by skip connection and incorporate it linearly.
- Estimation results show that the parameters for bicycling and walking are negative.
- The travel time, which is an important feature, is incorporated linearly, the accuracy of the model is inferior to normal NN(Dense NN **acc 0.945**).

Estimation of MNL

Estimated by MNL for comparison with NN and NN+MNL

	MNL:Generic parameter		**MNL**:Specified parameters	
Variable	params	t-value	params	t-value
ASC_train	-0.305	-0.887	-0.0285	-0.0743
ASC_bus	-1.10	-5.55 ***	-2.15	-7.41 ***
ASC_bike	-1.95	-8.82 ***	-1.86	-4.91 ***
ASC_cycle	1.29	7.97 ***	1.38	5.20 ***
ASC_walk	4.03	16.5 ***	3.66	12.7 ***
Cost	-0.0303	-0.725	-0.0160	-0.398
Short trip dummy	0.785	2.57 *	-0.0199	-0.0537
Access distance	0.670	9.43 ***	0.806	9.16 ***
Urban dummy	-1.99	-11.0 ***	-1.01	-3.81 ***
Travel time	**-0.166**	-15.8 ***		
Train travel time			**0.176**	28.5 ***
Bus travel time			**0.0187**	2.62 *
Car travel time			**0.0789**	14.6 **
Motorcycle travel time			**0.0230**	1.32
Bicycle travel time			**-0.0773**	-5.93 ***
Walk travel time			**-0.105**	-10.2 ***
Samples		1000		1000
Init log likelihood		-1779		-1779
Final log likelihood		-731		-695
Rho bar square		0.583		0.601
Accuracy		**0.761**		**0.817**

Generic parameter

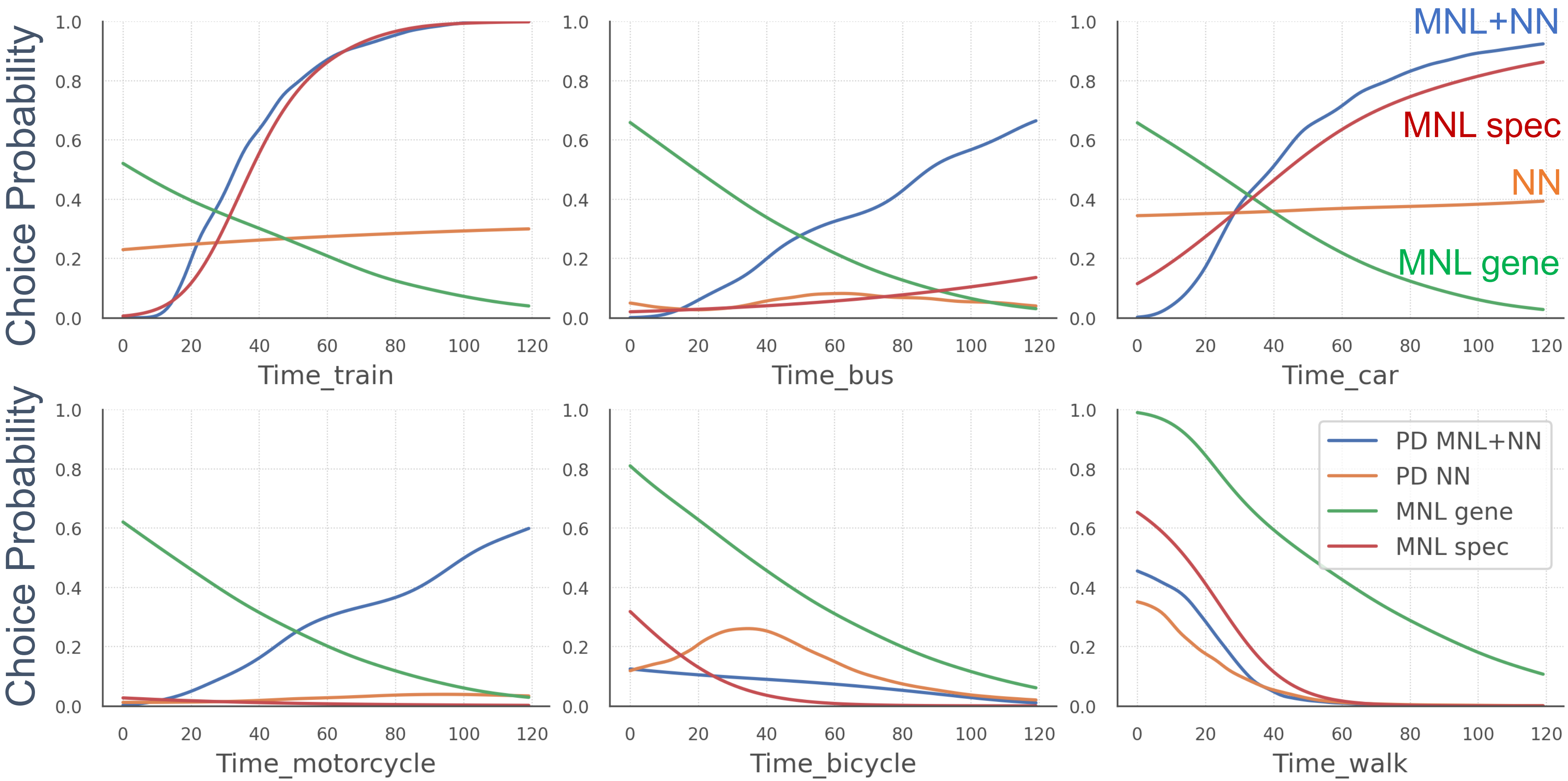
- The travel time parameter is negative and satisfies the sign condition

Specified parameters

- Compare with NN+MNL
- Parameters of bicycle and walk are considered negative because the mode avoids long distance travel.
- The positive/negative of the parameter is the same as the result of NN

Comparison of probability transition

PD(MNL+NN model, NN, Generic params MNL, Specified params MNL)



- ”MNL same” params are common, so it is similar outlines for all modes.
- “MNL different” and “MNL+NN” model have similar outlines because the signs of the parameters are the same in all modes.
- Although we do not know the true value, we believe that NN with nonlinearity can express the actual behavior well.

Comparison DCM(MNL) and NN

	DCM(MNL)		NN	DCM+NN
	generic	specified		
Interpre-tability	○		◎	◎
	• Feature parameters • **Sensitivity analysis** • **Elasticity** • Probability transition		• **Feature Importance** • Sensitivity analysis **(Nonlinear)** • Group **heterogeneity** • Feature **contribution** • Overall and **individual** parameters	• **Feature Importance** • Sensitivity analysis **(Linear)** • Group **heterogeneity** • Feature **contribution** • Overall and **individual** parameters
Accuracy	○	○	◎	○
	Acc 0.761 • Slow compute speed	**0.817**	• Acc **0.945** • High computation speed	• Acc **0.854** • High computation speed
Logicality	○		△	△
	• Utility Function		• **Minimize loss function**	• **Minimize loss function**

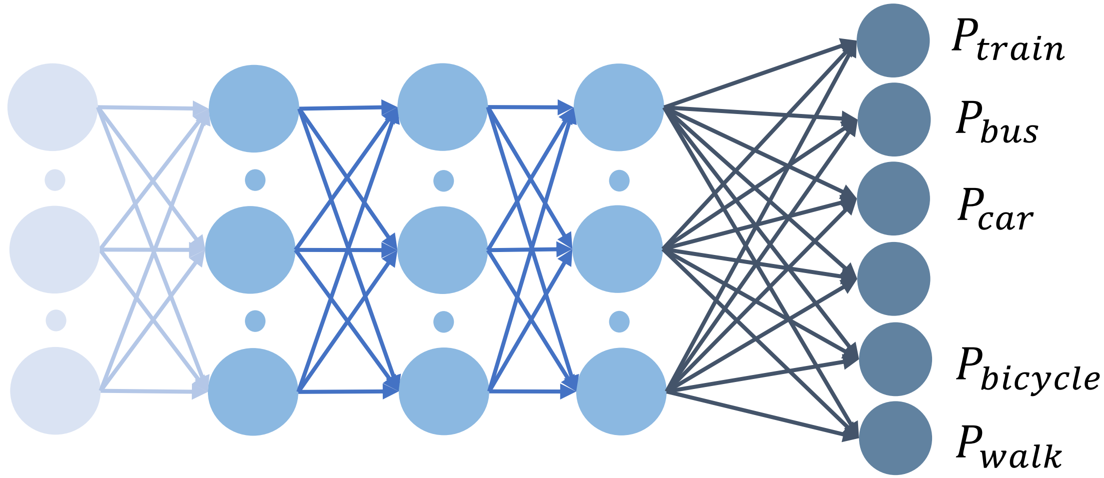
Use **ML** when **descriptive performance is more important than logicality**

Contents

1. Background and Objective of the Study
2. What is XAI(explainable AI)?
3. Application of XAI
4. Change Neural Network structure
   and comparison of MNL and NN
5. **Activity Simulation in Tokyo area**

Structure of NN based Activity model

Features



	Features from Tokyo PT data	Target val.
Activity	All Features	11 choice HWH, HSH, HPPH…
Dept. time	Without Dept. time, Arr. Time	17 choice 5:00, 6:00 … 21:00
Location	Without zone infomation	615 choice 千代田区, 港区, 江東区…
Mode	Without Main mode	6 choice Train, Bus, Car Motorcycle, Bike, Walk

Structure of NN based Activity model

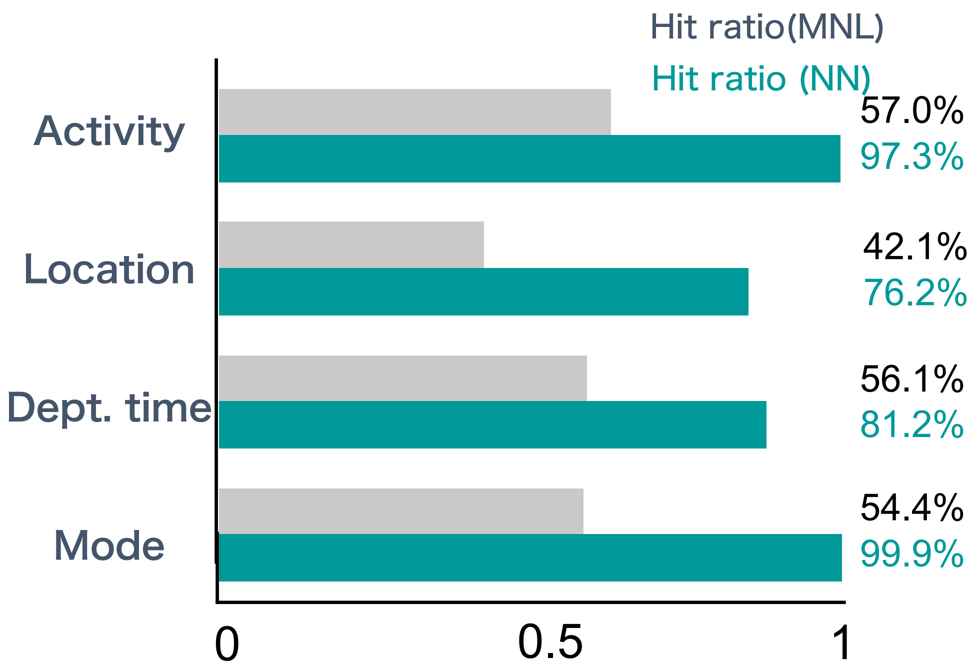
Goodness of fit

- NN has High hit ratio more than MNL.
- All models showed improved generalization performance.

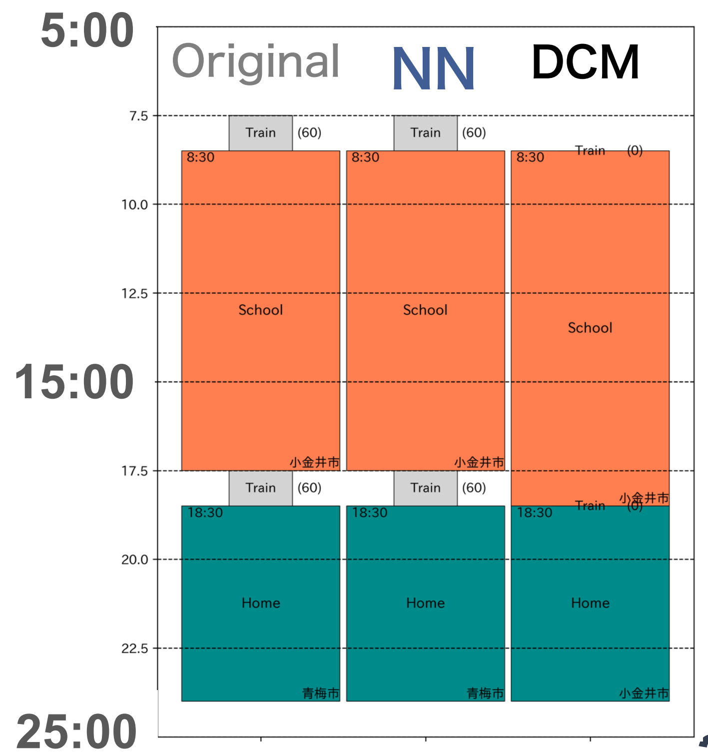
Forecasting

- More accurate prediction of activity than DCM

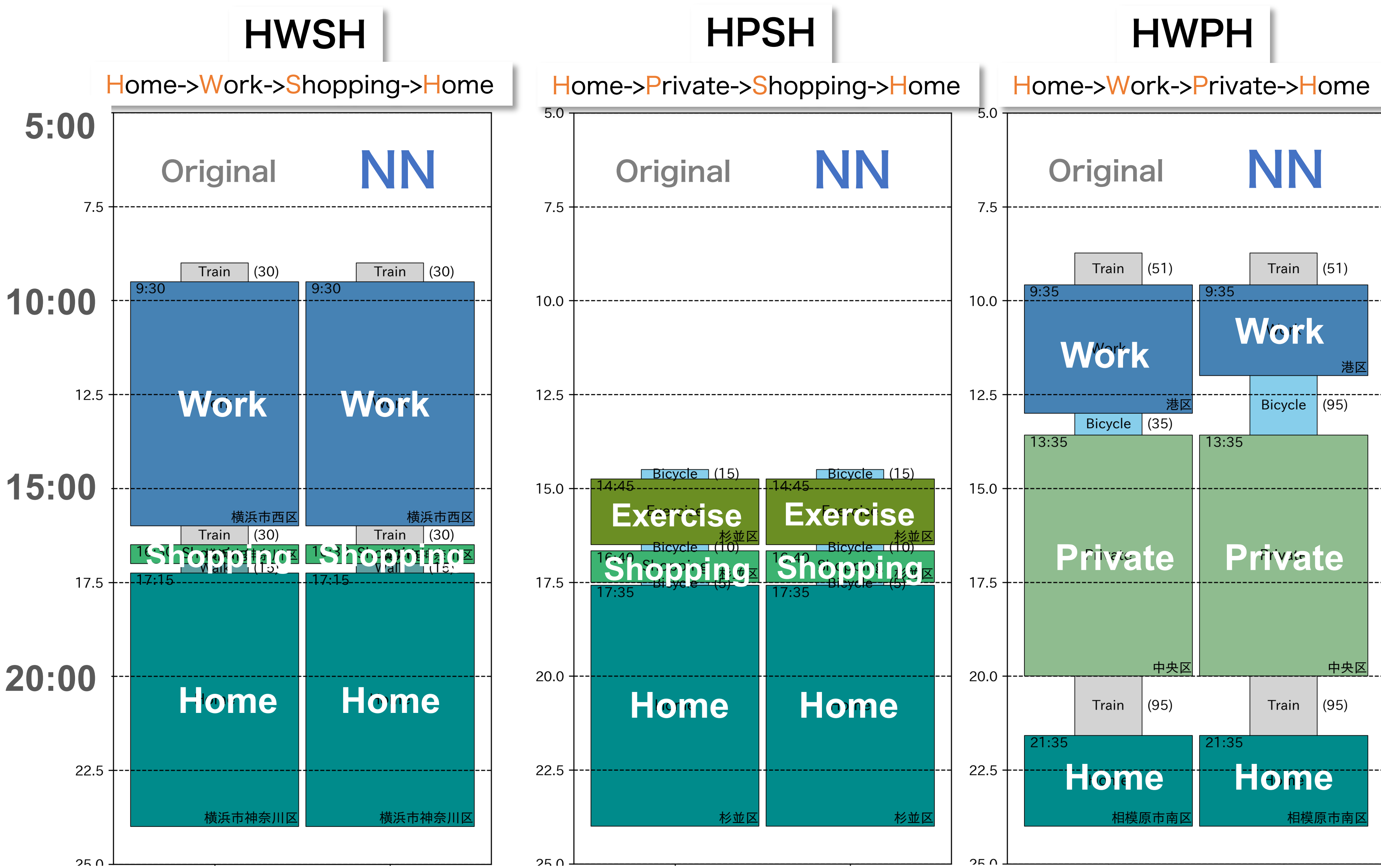
Hit Ratio



Activity within day



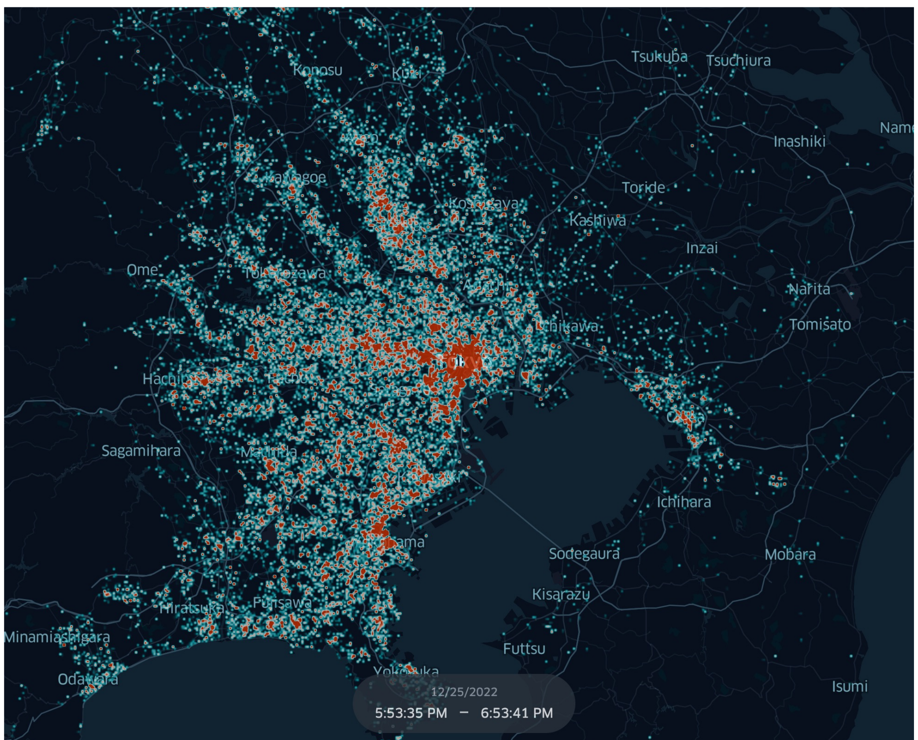
Example of prediction result



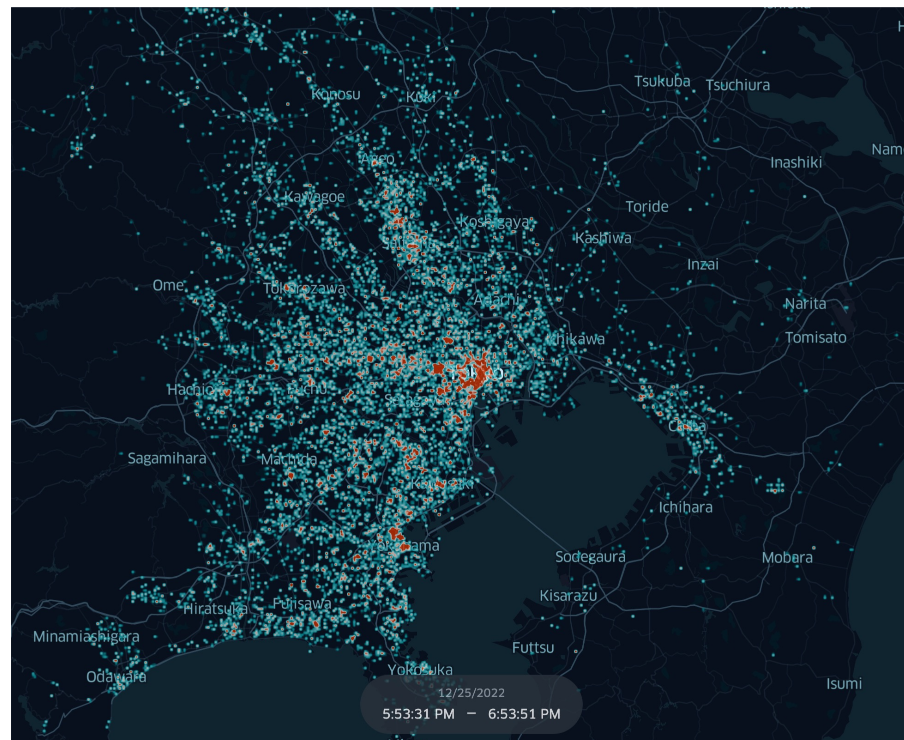
Simulation analysis of earthquake in Tokyo

- Setting hypothetical earthquake scenario.
- The earthquake was assumed to occur at 6:00pm.
- Visualization of 1 hour (6:00-pm7:00pm) trip “uncontrol” and “control” case from the time of the disaster.

uncontrolled



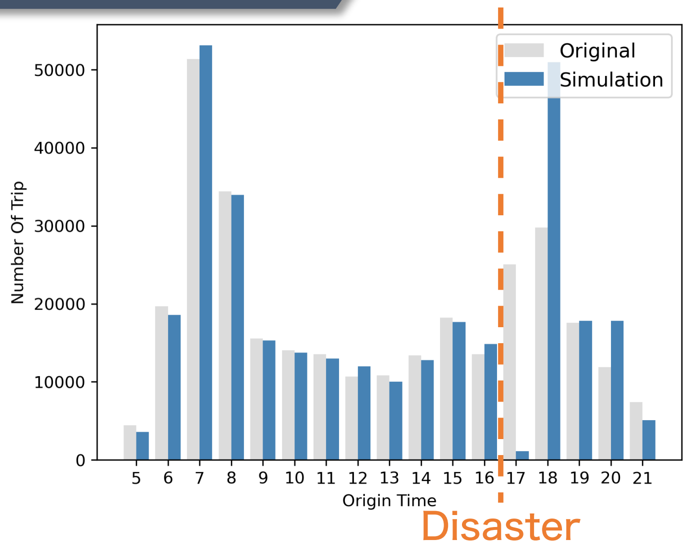
controlled

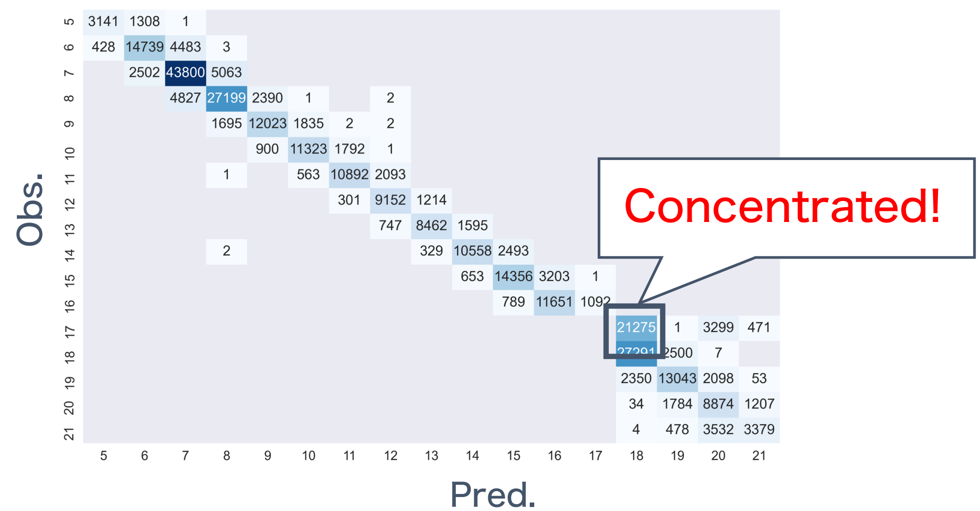


Structure of Activity model

uncontrolled

assume “freedom” departure time choice

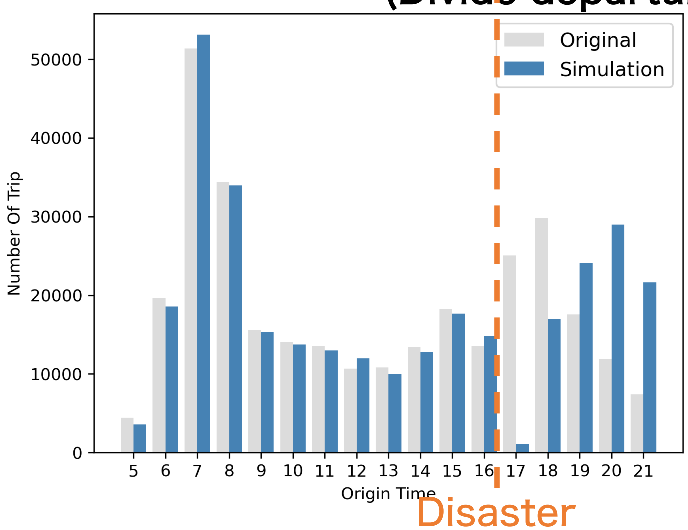


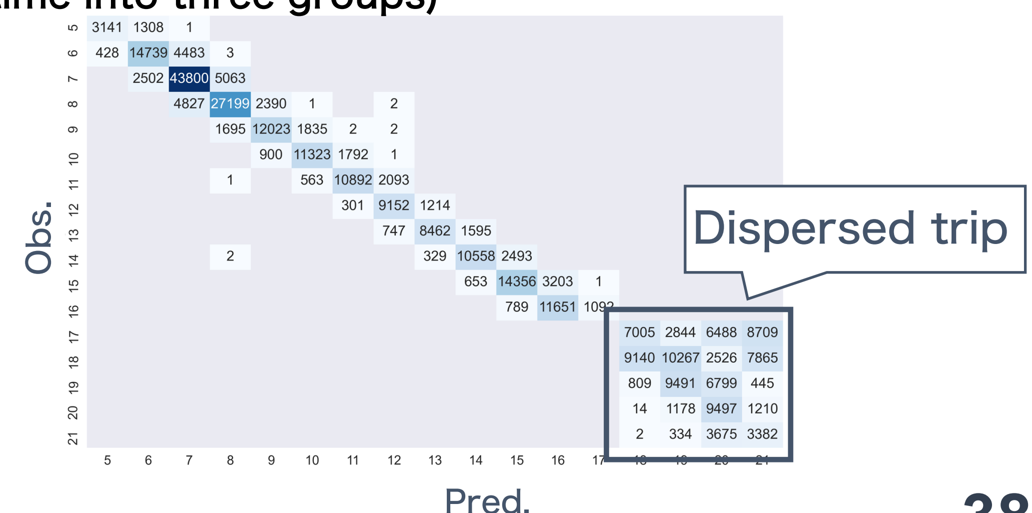


controlled

assume “control” departure time choice

(Divide departure time into three groups)





Thank you for your attention!

Spare Slide

Conclusion

- **Transportation choice model by machine learning**
  - Interpretability is taken into consideration by the **interpretation index (XAI) in machine learning**
    - PI,PD,CPD,SHAP
  - Ability to interpret models from both macro and micro perspectives
  - Elasticity can be reflected by changing the model structure by skip connection

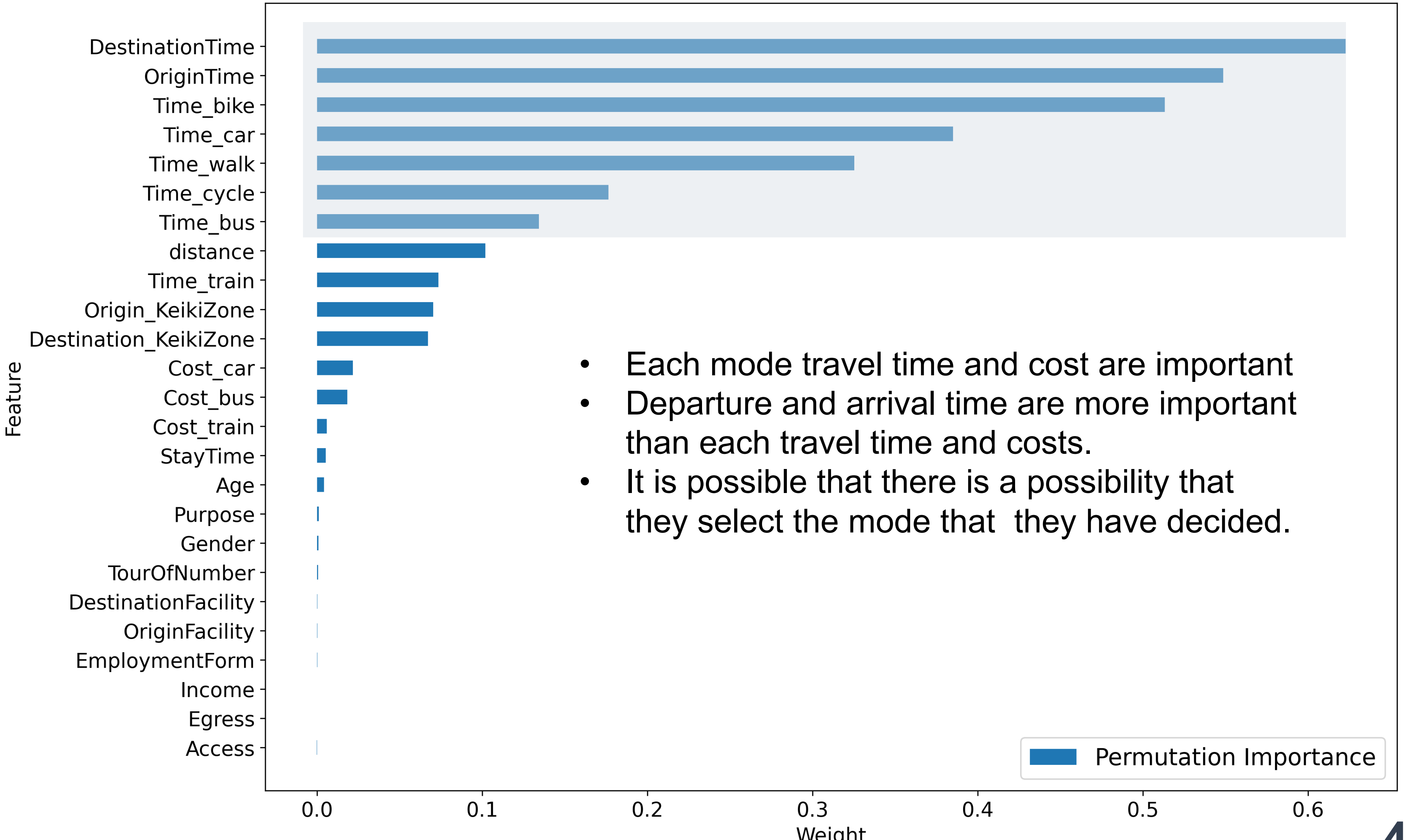
- **Discrete Choice Model and Machine Learning "Proper use" and "Combined use"**

  Proper use
  - Discrete choice model is used when theory is required such as decision of transportation policy
  - Use machine learning when descriptive performance is more important than theoretical

  Combined use
  - Using results derived from a theoretical model (discrete choice model) and a high-precision model (NN) makes it possible to make more accurate decisions.

Permutation Importance(PI)

- Applying PI to NN transportation mode choice model
  - ➢PI: calculate feature importance

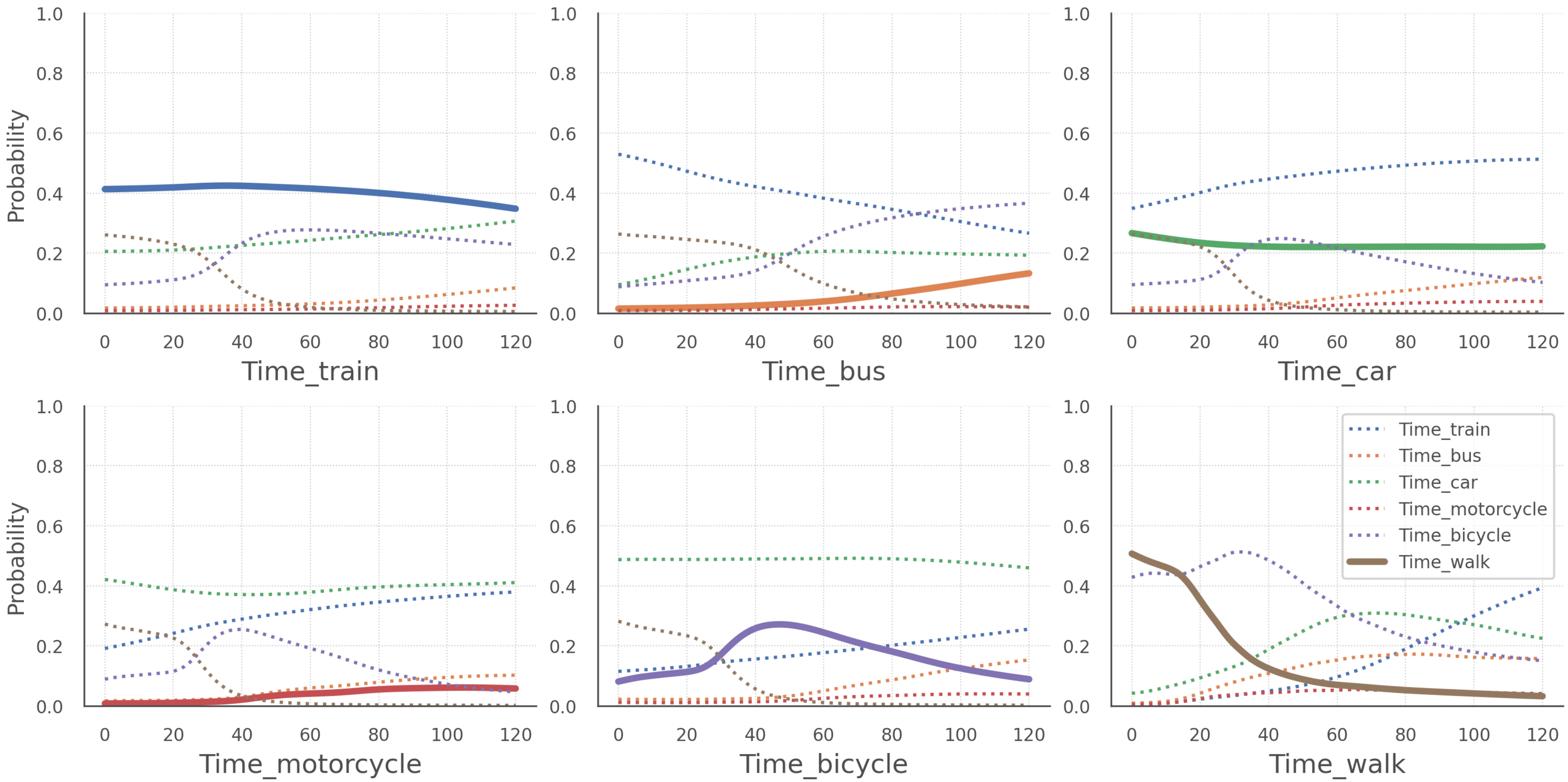


- Each mode travel time and cost are important
- Departure and arrival time are more important than each travel time and costs.
- It is possible that there is a possibility that they select the mode that they have decided.

Partial Dependence(PD)

Reflected the selection probability of other modes

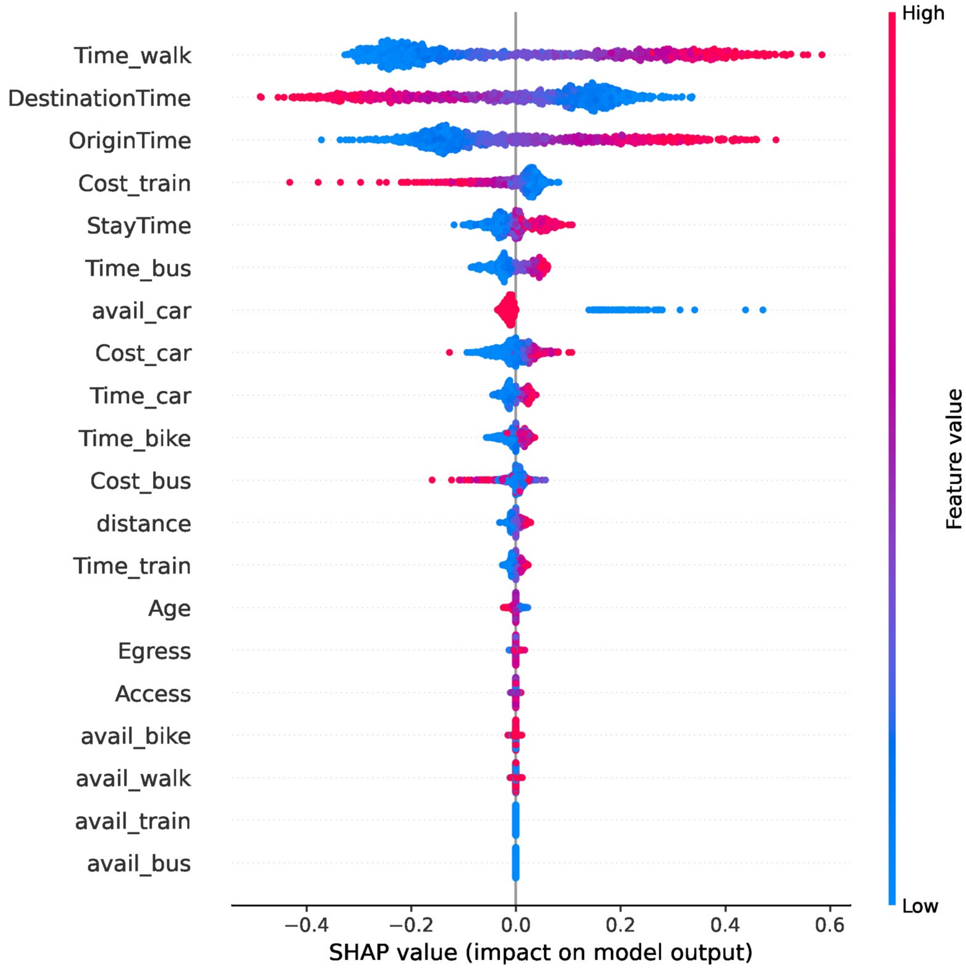
➢horizontal axis: **Travel time(minutes)**, vertical axis: **Probability**



- Add stochastic transitions of other modes in addition to the stochastic transitions of the target mode
  - Confirmed it, the nonlinearity of other modes was confirmed

beeswarm plot

➢In addition to the summary plot, a graph reflecting the size of the feature value



Feature Importance in railroad

- Larger feature values are redder, smaller feature values are bluer.
- The higher the feature value, the more important it is.

- For travel time
  The greater the travel time, the greater the SHAP value will also increase.

Greater probability of choosing rail

Permutation Importance(PI)

Importance of features by Permutation Importance

- Index for calculating feature importance in a model
- Shuffles certain features, resulting in worse prediction accuracy
  → Important features

random shuffle

PersonID	Age	Gender	Travel time
1	24	M	15
2	15	F	30
3	30	M	35
4	46	F	20

PersonID	Age	Gender	Travel time
1	24	M	30
2	15	F	20
3	30	M	15
4	46	F	35

- Can be calculated by performing the same process for each feature
- A feature is important in the model if the error in the model predicted by the randomized feature is large.
  - The difference of the loss function is the feature importance.
- Importance for model accuracy and causal interpretation is inappropriate
  - It is important to perform deeper analysis on features of higher importance

Partial Dependence(PD)

How each feature affects the model's predictions

- Calculate the average relationship between features and predictions
- It is possible to determine whether the model's predictions increase as the feature size increases, or whether the relationship is nonlinear
- Analyze sensitivity by moving only certain features while fixing other features

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

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

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

Averages forecast results

Shapley Additive exPlanations(SHAP)

- Reflecting **Shapley Value** in Cooperative Game Theory in Machine Learning
  - **Shapley Value**: Indicators used as criteria for profit sharing
  - Used to distribute the profits from the cooperation of multiple players according to each player's contribution.

participant	reward
A	6
B	4
C	2
A,B	20
A,C	15
B,C	10
A,B,C	24

Cooperative Game Theory

Replace each feature in the machine learning model

feature	Predicted Change
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

Machine Learning

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

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