

개인 맞춤형 대중교통 경로 추천을 위한 잠재특성 반영 연구

연세대학교 도시공학과 도시공학종합설계
2020143517 최윤서, 2020143530 김지원 지도교수 김진희

1. 연구 배경 및 목적

- 최근 모빌리티의 발전 중 가장 큰 변화는 인프라 중심에서 이용자 중심으로 변화하고 있다는 것이고, 이용자 다양성, 개인화에 대한 고려는 스마트 모빌리티에 대한 선택모형에서 필요한 요소임

- 같은 출발지에서 도착지까지 대중교통이라는 동일한 수단을 이용하는 통행이라더라도 길찾기 서비스에서는 다양한 경로가 제공되고 사람들은 자신의 잠재특성에 따라 각자 최적이라고 느끼는 서로 다른 경로를 선택함

- 본 연구에서 정의하는 ‘**잠재특성**’이란?

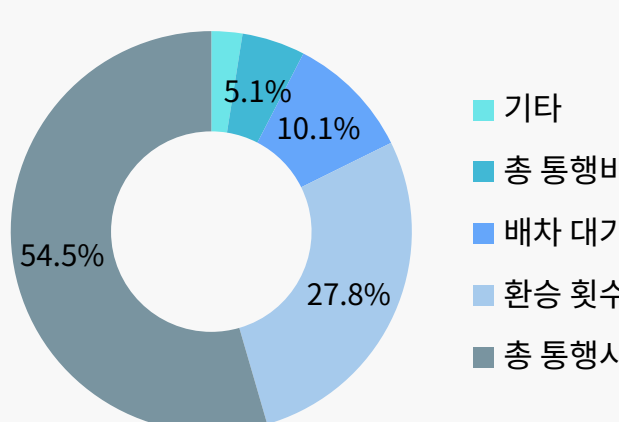
- 대중교통 경로를 선택할 때 개인에 따라 달라지는 통행속성에 대한 민감도를 의미함
- (예시) 도보 접근 시간에 민감한 A는 환승을 감수하더라도 3번째 경로를 선호하고, 환승 대기 시간에 민감한 B는 도보 접근 시간이 길더라도 환승이 없는 1번째 경로를 선호하는 성향이 내재되어 있음

- 설문조사 결과 대중교통 이용시 중요한 요소는 개인에 따라 다양하게 나타났으며, 대중교통 길찾기 서비스에서 ‘추천경로가 개인의 선호나 습관을 반영하지 않는다(38.6%)’는 점에서 가장 불편함을 느꼈다고 응답함. 이에 대중교통 길찾기 서비스에서 개인화된 접근의 필요성이 대두됨

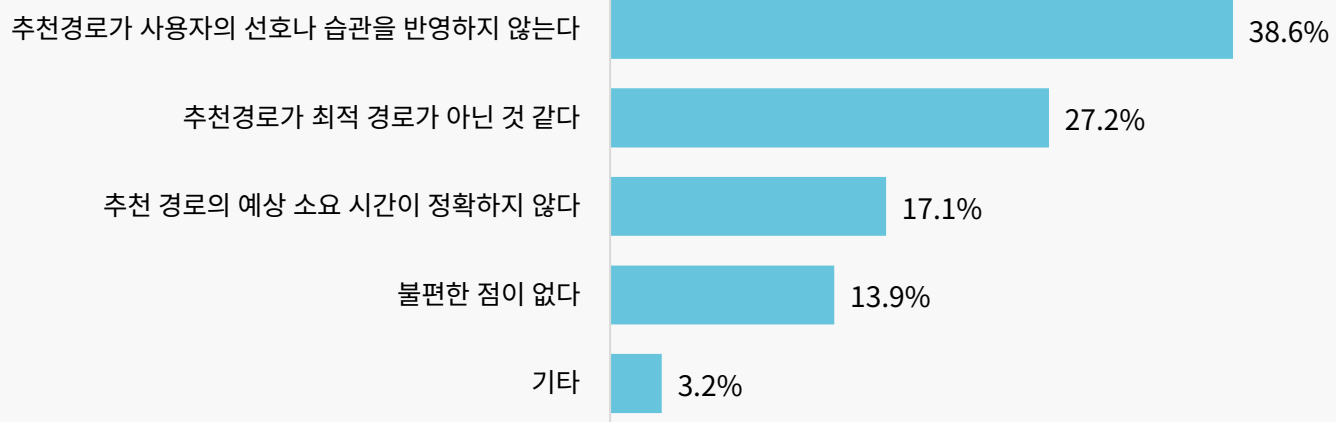


대중교통 길찾기 경로 추천 예시

대중교통 이용시 가장 중요한 요소



대중교통 길찾기 서비스에서 가장 불편했던 점



- 본 연구에서는 사람들의 평균적인 수준을 추정하여 교통 서비스를 제공한 기존의 선택 모형과 다르게, 잠재선호(SP) 조사를 시행하여 대중교통 경로 선택에 미치는 요인을 파악하고 개인에 따라 달라지는 잠재특성을 고려한 교통서비스 제공(대중교통 경로 추천 방안)을 제안하고자 함

- 효용함수란?

- 효용은 선택가능한 대안 (본 연구에서는 대중교통 경로)에 대한 매력도의 크기를 의미함
- 통행속성들을 반영하여 하나의 스칼라 값으로 환산할 수 있는 효용함수를 구성하고, 이때 개인은 효용이 최대가 되는 경로를 선택하게 됨

$$U_{in} = \beta^T X_{in} + \epsilon_{in} \quad \text{choose } U_{kn} \text{ such that } U_{kn} = \arg \max U_{in} \\ (U = \text{효용}, X = \text{속성벡터}, \epsilon = \text{error}, i = \text{선택가능한 모든 경로}, n = \text{사람}, k = \text{선택한 경로})$$

- Logit model은 오차항 ϵ_{in} 이 독립적으로 동일한 Gumbel 분포를 따른다고 가정하고, 개인 n이 경로 j를 선택할 확률은 다음과 같이 표현함

$$P_{jn} = \frac{\exp(\beta^T X_{jn})}{\sum_i \exp(\beta^T X_{in})} \quad \forall j$$

2. 실험 방법

- 표본 수집 및 설문조사 설계

- 2024년 10월 7일 ~ 12월 6일간 수도권 거주자 학생 및 직장인 204명을 대상으로 온라인 설문을 시행함
- 표본의 사회경제적 속성으로 응답자의 54%가 남성으로 더 많은 비중을 차지했으며, 20대가 61%로 가장 많았고, 40대가 31%, 30대가 8%로 나타남.

- 잠재선호(Stated Preference) 선택실험 설계

- 출근 또는 등교의 목적으로 연세대학교 제1공학관에서 강남역 11번 출구까지 이동할 때 대중교통을 이용하는 상황에서 서로 다른 통행정보 속성으로 구성된 두 개의 경로 A,B 중에 선택하도록 문항을 설계함
- 응답자에게는 총 8번의 경로선택상황이 주어졌으며(Random Block), 8번의 선택에 내재된 응답자별 잠재특성을 파악하고자 함
- Orthogonal Design 기법으로 총 32개 조합으로 속성 수준을 할당하고 4개의 Block으로 분리

잠재선호조사를 위한 속성값 레벨

Attribute	Definition	Attribute level
통행 요금	승차부터 하차까지 소요되는 모든 비용의 총합 (원)	1500원, 1750원, 2000원, 2250원
차내 이동 시간	대중교통 수단 내에서 이동하는 실제 소요 시간의 총합 (분)	35분, 40분, 45분, 50분
도보 접근 시간	승차 전 정류장까지 걷는 시간과 하차 후 목적지까지 걷는 시간의 총합 (분)	10분, 12분, 14분, 16분
환승 및 배차 대기 시간	환승 시 소요되는 시간과 정류장에서 대기하는 시간의 총합 (분)	6분, 8분, 10분, 12분

- Proposed Model

개인의 선택이 누적될 때마다 **개인에 맞는 잠재특성 벡터**를 추정하는 **Bayesian Updating Logit Model**을 제안

- β 를 상수가 아닌 확률변수로 가정하고, 효용함수는 다음과 같이 정의한다

$$U_{in} = E(\beta_{nk})^T X_{in} + \epsilon_{in} \quad , \beta_{nk} \sim \text{Normal distribution}, \epsilon_{in} \sim \text{iid Gumbel distribution}$$

- β_{nk} 는 초깃값 β_0 로부터 Bayesian Updating 된다

$$\beta_0 \xrightarrow{\text{update with } Y_{n1}} \beta_{n1} \xrightarrow{\text{update with } Y_{n2}} \beta_{n2} \cdots \xrightarrow{\text{update with } Y_{n(k-1)}} \beta_{n(k-1)} \xrightarrow{\text{update with } Y_{nk}} \beta_{nk}$$

$$\begin{aligned} \text{1st update: } P(\beta_{n1} | Y_{n1}) &\propto P(\beta_0) \times P(Y_{n1} | \beta_0) \\ \text{kth update: } P(\beta_{nk} | Y_{nk}) &\propto P(\beta_{n(k-1)}) \times P(Y_{nk} | \beta_{n(k-1)}) \quad \forall k \geq 2 \end{aligned}$$

(Y_{nk} 는 개인 n의 k번째 선택, β_{nk} 는 개인 n의 k번째 선택까지 반영하여 업데이트된 잠재특성 벡터)

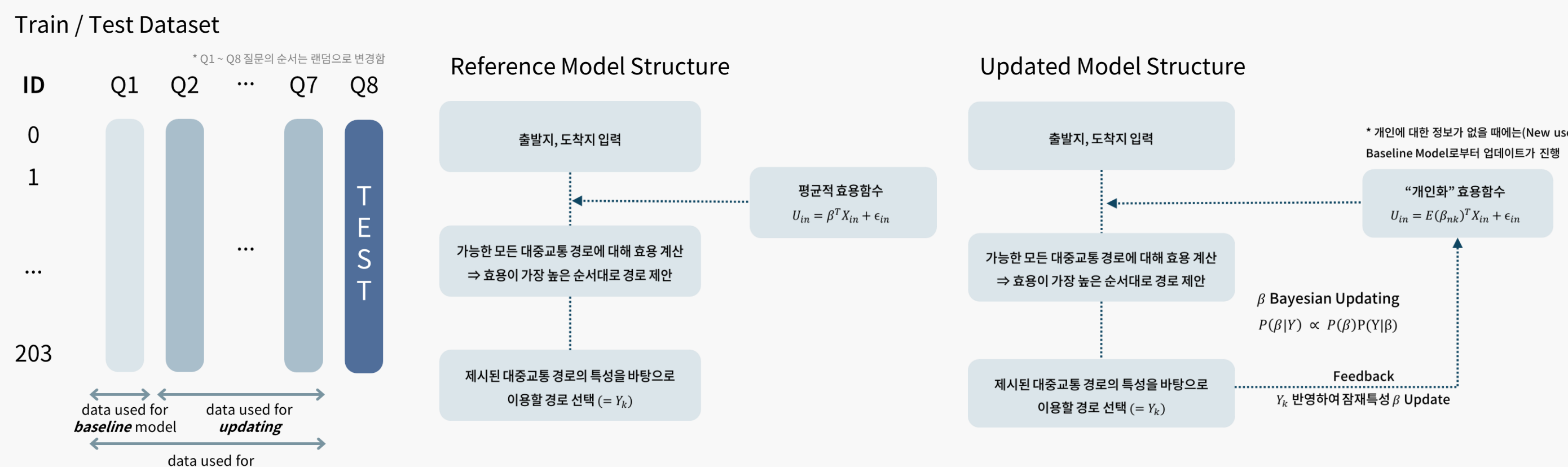
- $\bar{\beta}_0$ 는 Logit Model에서 추정되는 $\hat{\beta}$ 이고, β_0 의 분포는 다음과 같이 나타남다 $\beta_0 \sim N(\hat{\beta}, n \times \text{Var}(\hat{\beta}))$

$$(\because \beta_0 \sim N(\mu, \sigma^2) \Rightarrow \bar{\beta}_0 \sim N(\mu, \frac{\sigma^2}{n}) \text{ by Central Limit Theorem})$$

- Dataset Split & Model Definition

- 잠재선호 선택실험에 응답한 204명 각 사람에 대하여 8개의 응답 중 하나를 무작위로 선택하여 Test dataset으로 설정하고, 나머지 7개의 응답을 Train dataset으로 사용
- Reference Model과 Updated Model을 각각 추정하여 Test dataset에서 개인의 선택 결과를 잘 예측하는지 비교 분석하고자 함

Model Name	Definition	Description
Reference Model	Traditional Logit Model	<ul style="list-style-type: none">모든 데이터 (204×7개)를 기반으로 MLE 추정모든 사람들을 평균적으로 가장 잘 설명하는 모형
Baseline Model	Bayesian Updating Logit Model with β_0	<ul style="list-style-type: none">초기의 데이터 (204개)를 기반으로 MLE추정모든 사람에 대해 평균적인 민감도 $E(\beta_0)$를 계수로 가지는 모형
Updated Model	Bayesian Updating Logit Model with β_{nk}	<ul style="list-style-type: none">개인(n)의 누적된 선택(k)을 반영하여 Baseline model에서 업데이트가 이루어진 $E(\beta_{nk})$를 계수로 가지는 모형



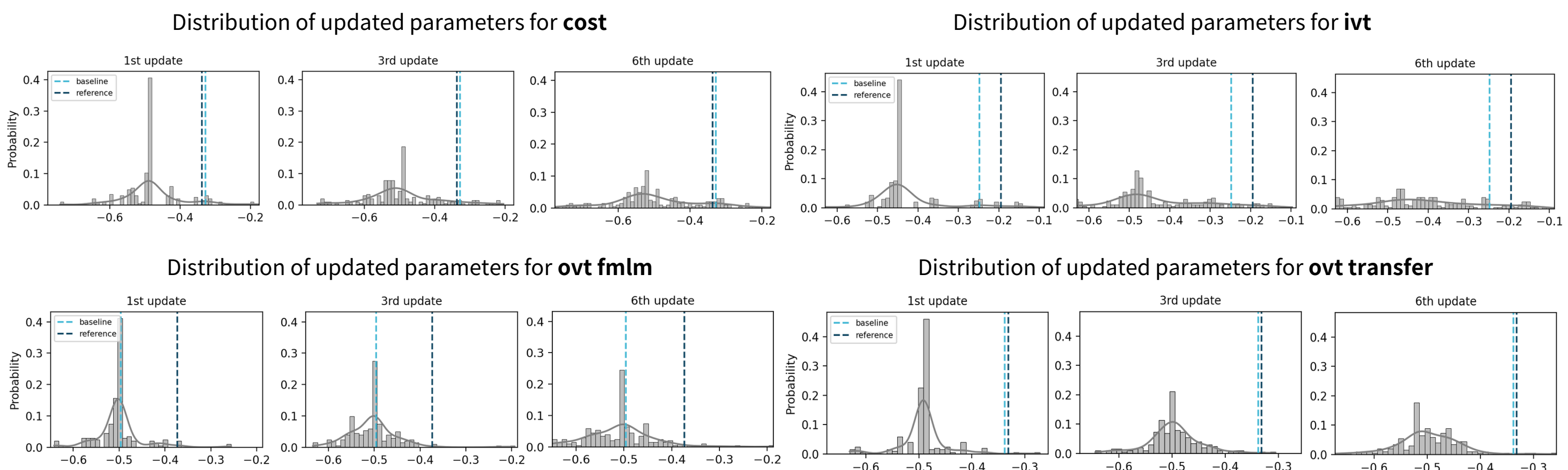
3. 실험 결과

- 추정결과

Reference Model					Baseline Model				
variables	coefficient	std err	Z	p-value	variables	coefficient	std err	Z	p-value
cost diff	-0.3377	0.023	-14.740	0.000 ***	cost diff	-0.3279	0.063	-5.169	0.000 ***
ivt diff	-0.1950	0.013	-15.080	0.000 ***	ivt diff	-0.2479	0.041	-6.040	0.000 ***
ovt fmlm diff	-0.3732	0.027	-13.893	0.000 ***	ovt fmlm diff	-0.4960	0.082	-6.029	0.000 ***
ovt transfer diff	-0.3316	0.028	-11.762	0.000 ***	ovt transfer diff	-0.3379	0.084	-4.025	0.000 ***
Number of observations: 1428 / Log-Likelihood: -574.91 ***: Significance level < 0.01					Number of observations: 204 / Log-Likelihood: -75.570 ***: Significance level < 0.01				

- Updated Model

- Baseline Model로부터 Update stage(1~6)에 따라 전체 표본(204명)에 대하여 각 통행속성별 파라미터가 어떻게 변하는지 관찰
- 모든 통행속성에 대하여 초기에는 특정 값 근처에 집중된 분포를 보이지만, update가 계속됨에 따라 σ 가 점점 커짐
- 이는 개인의 선택에 내재된 잠재특성이 반영되면서 개인에 맞는 추정치로 분산됨을 의미
- 아래 표는 통행속성별로 1st update 후, 3rd update 후, 6th update 후 표본의 파라미터 분포를 나타냄

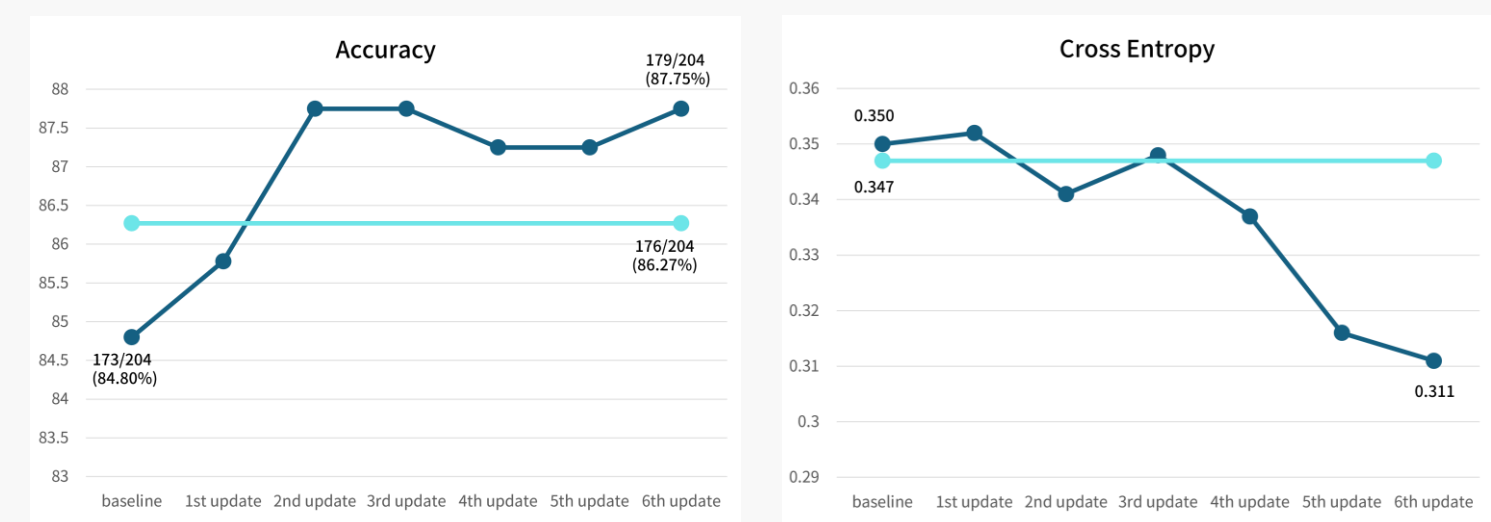


- Reference Model과 Updated Model 성능 비교

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N 1(\hat{y}_i = y_i)$$

$$\text{Cross Entropy} = \frac{1}{N} \sum_{i=1}^N -[y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i)]$$

* Cross Entropy란?
모델의 예측 확률 분포가 실제 정답과 얼마나 가까운지를 측정하는 지표로, 작을수록 좋음. 예측의 맞고 틀림을 넘어서, 모델이 얼마나 더 높은 신뢰도로 예측했는지 비교가능함



4. 결과 해석

- 확률론적 해석

- Updated Model이 어떻게 더 예측을 잘 할 수 있었는지 해석하기 위해서 test data 중 랜덤으로 질문 1개를 뽑아 (Block 4의 Q3: 우측 표) 확률론적 해석을 진행
- Block 4 응답자 중 Reference & Updated Model 모두 예측에 성공한 ID 181, Updated Model만 예측에 성공한 ID 111, 모두 예측에 실패한 ID 193을 선정하여 예측된 확률을 비교

통행속성	경로 A	경로 B
통행요금	2000원	1500원
차내 이동 시간	40분	40분
도보 접근 시간	14분	16분
환승 및 배차 대기 시간	6분	8분

Reference & Updated Model 비교 방법

경로 A,B 선택 확률 계산 결과

- (1) 효용함수 적용

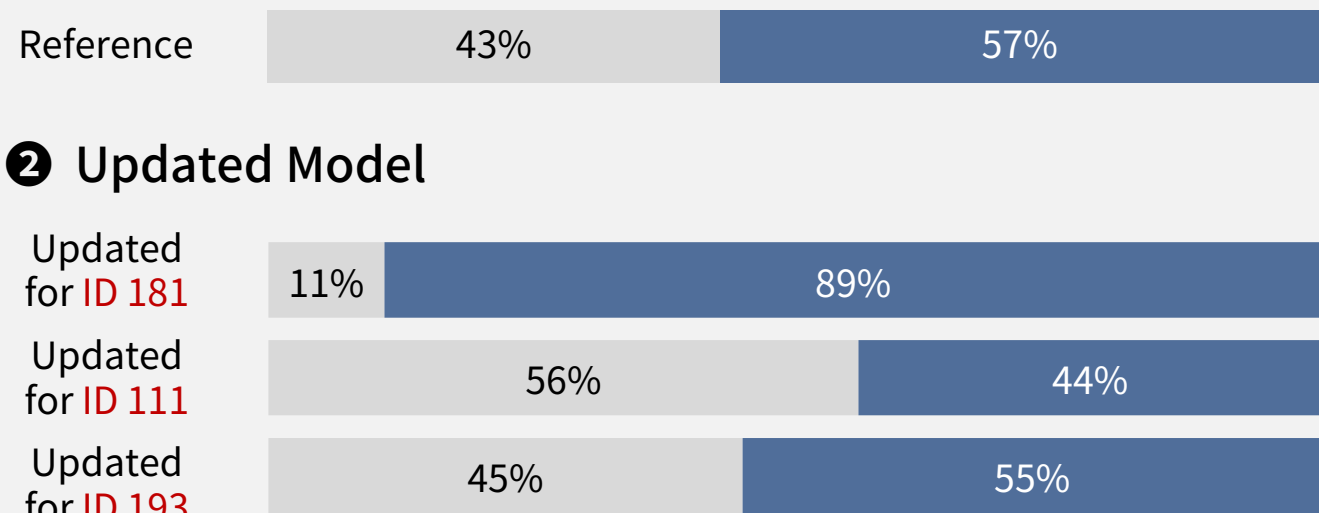
- ① Reference Model

- 모든 사람에게 동일한 평균적 효용함수 적용

- ② Updated Model

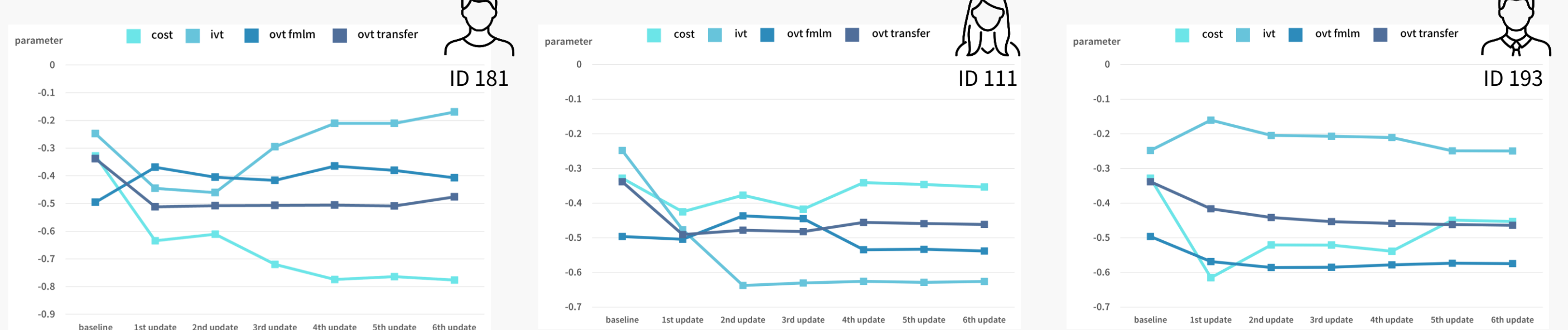
- 개별 잠재특성을 반영하여 파라미터가 업데이트 된 개인화된 효용함수 적용
- 따라서 모델은 주어진 경로 선택 상황에서 개인마다 어떤 경로를 어떤 확률로 선택할 것인지 서로 다른 값으로 예측을 수행

- ① Reference Model



ID	Answer	Reference Prediction	Updated Prediction	해석
181	B	B (57%)	B (89%)	둘 다 정답을 예측했지만 updated model의 정답 예측 확률 증가
111	A	B (57%)	A (56%)	정답을 맞춘 확률이 점차 증가하여 updated model만 예측에 성공
193	A	B (57%)	B (55%)	둘 다 오답을 예측했지만 updated model의 오답 예측 확률 감소

- 효용함수 관점의 해석



위와 같이 ID별로 속성별 파라미터가 stage에 따라 update됨을 확인할 수 있으며, 이를 이용하여 경로선택 효용 비교를 하면 아래 표와 같음

Block 4 - Q3		Reference Model		Updated for ID 181		Updated for ID 111		Updated for ID 193	
variables	compare	β	Utility Score	β_{181}	Utility Score	β_{111}	Utility Score	β_{193}	Utility Score
통행요금 (100원)	B가 A보다 500원 저렴	-0.3377	B + 1.69	-0.7763	B + 3.88	-0.3534	B + 1.77	-0.4528	B + 2.26
차내 이동시간	-	-0.1950	-	-0.1697	-	-0.6260	-	-0.2492	-
도보 접근시간	A가 B보다 2분 짧음	-0.3732	A + 0.75	-0.4074	A + 0.81	-0.5380	A + 1.08	-0.5744	A + 1.15
환승 및 배차 대기시간	A가 B보다 2분 짧음	-0.3316	A + 0.66	-0.4760	A + 0.95	-0.4613	A + 0.92	-0.4638	A + 0.93
(SUM)		A + 1.41 / B + 1.69		A + 1.76 / B + 3.88		A + 2.00 / B + 1.77		A + 2.08 / B + 2.26	
해석		항상 경로B로 예측		요금에 더 민감한 잠재특성 반영되어 더 높은 확률로 B를 선택할 것으로 예측		도보접근, 환승및배차대기시간에 더 민감한 잠재특성 반영되어 A를 선택할 것으로 예측		Reference model과 같이 오답이지만, 잠재특성 반영되어 두 경로 사이의 효용차이가 더 적어짐	

5. 결론

- 실제 적용 예시

- 본 연구에서 제안한 Bayesian Updating Logit Model을 이용하여 계산된 개인화된 효용함수를 적용한 결과, 동일한 출발지와 목적지에 대하여 각 ID별 추천 경로 순서가 바뀜
- 비용과 환승 및 배차 대기 시간에 민감한 ID 181은 환승이 없는 경로가 추천되고, 비용에는 덜 민감하지만 차내이동시간에 매우 민감한 ID 111은 조금 비싸더라도 빠르게 갈 수 있는 경로부터 우선 제시됨
- 계수의 변화가 크게 나타나지 않은 ID 191은 Reference Model의 추천과 동일함
- 개인의 누적된 선택으로부터 잠재특성을 자동으로 추출해내 업데이트함으로써 모두에게 평균적으로 좋은 동일 추천경로가 아닌, 사용자 개인에 맞는 경로 추천이 가능함

- 의의 및 한계

- 대중교통 경로 추천에 개인의 잠재특성을 반영함으로써 이용자 중심의 서비스 제공이 가능해짐
- 개인의 선택 경험이 계속 누적되면서 실시간으로 업데이트가 가능하고, 시간의 흐름에 따라 더욱 정교하고 개인에게 적합한 추천이 가능해짐
- 만약 하나의 경로가 다른 경로에 비해 우세한 상황에서는 개인화된 추천의 효과가 제한적일 수 있음
- Updated model의 제안 경로에 대하여 해석이 불가능한 경우가 있었으며, 데이터의 수를 늘리면 해석 가능성에서의 개선을 기대할 수 있음

