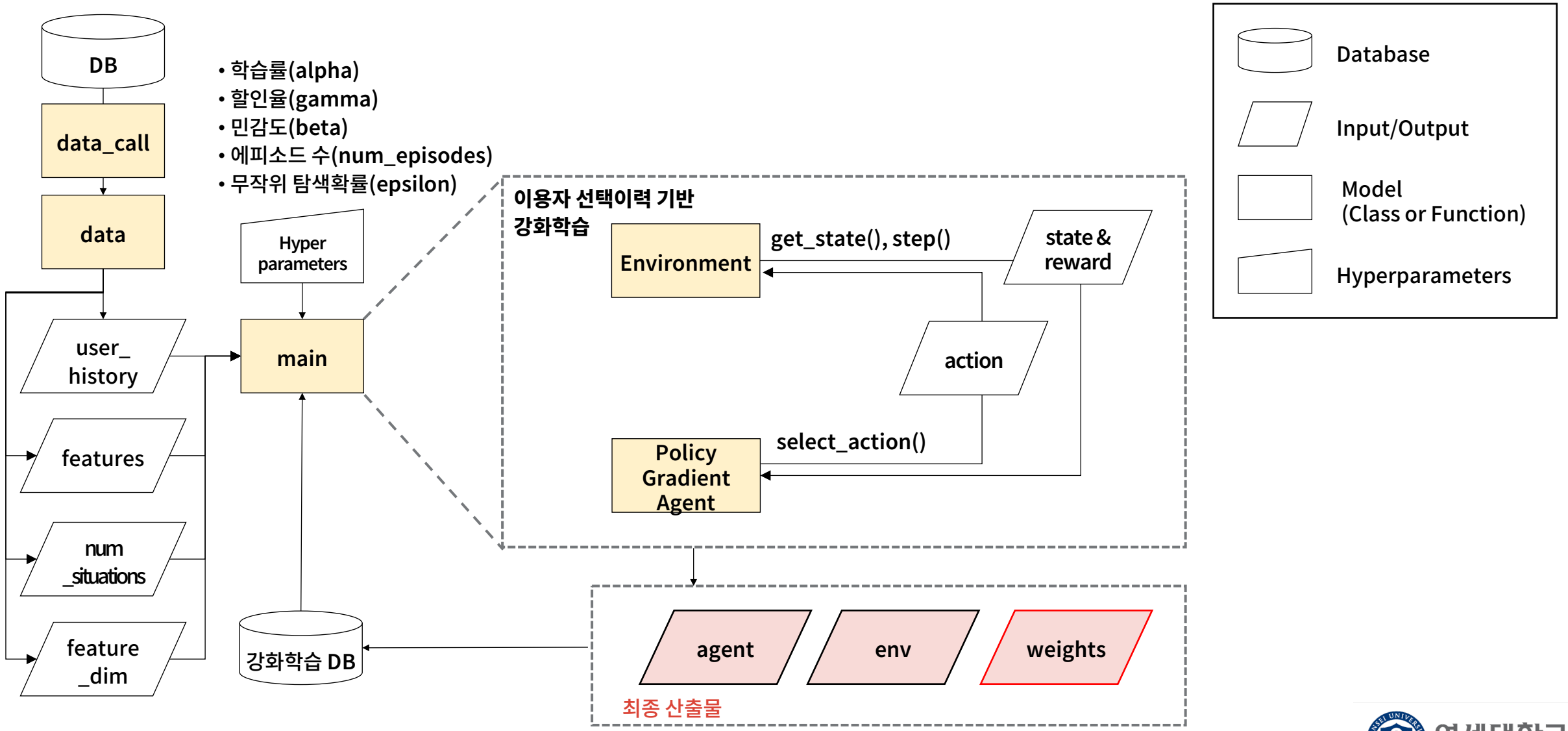
자율주행 수요응답 대중교통 이용자 이용패턴 및 선호경로 분석

연세대학교
YONSEI UNIVERSITY

Environments

- **Python 3.11.1**
- **Package List**
  - pandas (version 2.2.2)
  - numpy (version 2.0.0)
  - random(파이썬 내장 패키지)
  - pickle (파이썬 내장 패키지)

Framework 자료 호출 및 전처리



연세대학교 YONSEI UNIVERSITY

data_call & data

- **Database에서 이용자 ID를 기준으로 이력 데이터를 불러오는 모듈 (※ 개별 테스트 미시행)**
- **이용자 수락 DB(user_history_df)와 배차 시 제공 정보 DB(features_df)에서 이용자 ID(id)를 입력하여 데이터를 불러옴**
  - 본 학습에서 사용한 user_history_df와 features_df는 리빙랩 대상지 500명을 대상으로 설문조사를 수행한 자료로 서비스 선택 문항별 이용자의 대안선택 이력과 문항에서 제시된 대안 설명변수를 포함하고 있음*(리빙랩 수락률 모형 설문조사 설명 참조)

data_call

구분	자료명	자료형태	설명
Input	path	[Int] (ID)	이용자 ID – DB환경에 따라 변경가능
Output	user_history_df(DB)	[DataFrame] (예시자료: user_history.csv)	이용자 ID(id), 이력순서(situation), 배차 후 이용자가 선택한 대안(choice)
	features_df(DB)	[DataFrame] (예시자료: features.csv)	이용자 ID(id), 이력순서(situation), 운영자 제공 대안(alternative; 2 = 거절을 의미), 이력정보(access, wait, ivt, egress, constant, Linc, license; 거절 대안은 constant를 제외한 모든 정보가 0)

data

구분	자료명	자료형태	설명
Input	id	[Int] (ID)	이용자 ID – DB환경에 따라 변경가능
	path	[DataFrame] (예시자료: user_history.csv)	이용자 ID(id), 이력순서(situation), 배차 후 이용자가 선택한 대안(choice)
	user	[Dictionary]	이용자ID를 키값으로 user_history_df와 features_df로 구성된 딕셔너리
Output	user_history	[List]	이력순서에 따른 이용자가 선택한 대안
	features	[Array]	이력순서에 따른 대안별 이력 정보
	num_situations	[Int]	배차 이력 횟수
	feature_dim	[Int]	배차 이력 정보 갯수

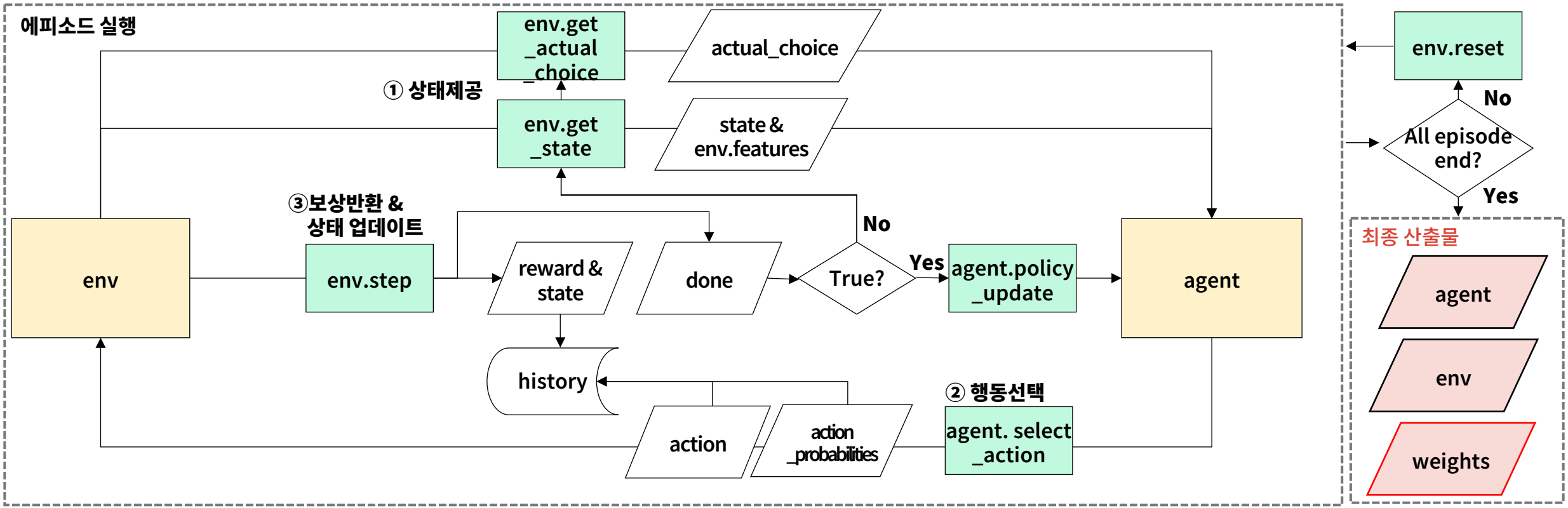
main[1/2]

- **호출한 데이터를 기반으로 강화학습을 수행하는 함수**
  - 이용자 ID 별 Environment(env)와 PolicyGradientAgent(agent) 객체 생성
  - 임의로 선정한 하이퍼 파라미터를 기반으로 학습수행
  - 각 에피소드의 총 보상을 기록하여 강화학습 수렴정도 파악

구분	자료명	자료형태	설명
Input	user_history	[List]	이력순서에 따른 이용자가 선택한 대안
	features	[Array]	이력순서에 따른 대안별 이력 정보
	feature_dim	[Int]	배차 이력 정보 갯수
	learned_weights	[Array] 1Xfeature_dim	(가중치 갱신 시) 이전에 학습된 Agent의 정책함수 가중치를 적용
Output	agent(trained_agent)	[Instance]	학습된 강화학습 Agent
	env(trained_env)	[Instance]	강화학습시 활용한 Environment
	agent.weights(weights)	[Array] 1Xfeature_dim	학습된 강화학습 Agent의 정책함수 가중치

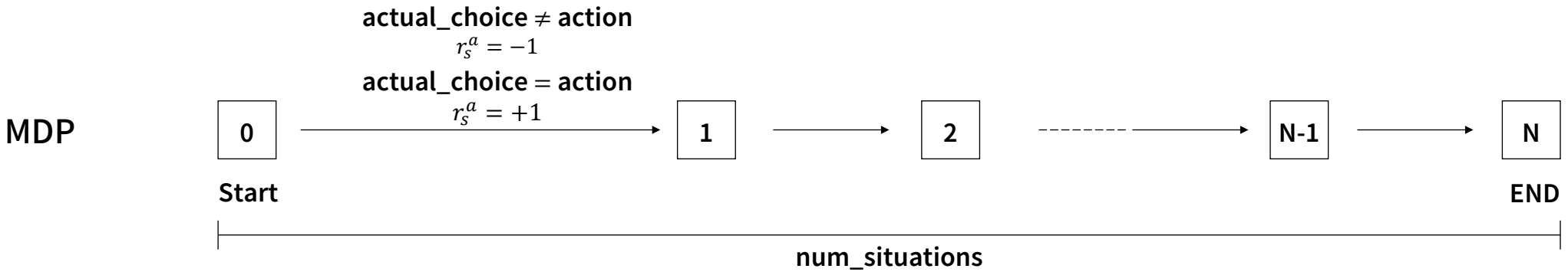
main[2/2]

- 에피소드 종료 시점에 도달할 때 까지 함수 내 로컬 인스턴스 env와 agent가 상태제공→행동선택→보상반환 & 상태 업데이트 → 상태제공 구조를 반복함
- 에피소드가 종료되면 agent는 policy_update 과정을 통해 정책을 업데이트하며, env는 다음 에피소드 수행을 위하여 상태 초기화
- 구조)



Environment

- 강화학습을 위한 MDP와 Reward 환경
- 학습해야하는 State 환경(MDP)를 제공하고 PolicyGradientAgent가 Action을 취하면 그에 따라 Reward를 부여
  - Reward는 Agent가 주어진 State에서 실제 이력 내역을 맞춘 경우 +1, 틀린경우 패널티 -1을 부여하는 형태
  - 마지막 배차 수력 이력에 도달할 때까지 에피소드 진행(Action에 따른 전이 확률$P_{ss'}^{a}$은 1 무조건 다음 state로 진행)



구분	자료명	자료형태	설명
Input	user_history	[List]	이력순서에 따른 이용자가 선택한 대안
	features	[Array]	이력순서에 따른 대안별 이력 정보
Output	actual_choice	[List]	이용자 호출 이력 순서에 따른 대안 선택 결과
	state	[Int]	MDP에서 Agent의 상태위치
	reward	[Int]	Agent의 action에 따른 리워드
	done	[Int]	에피소드 종료 여부, 종료 시 마지막 state 번호 반환

PolicyGradientAgent

- **Agent의 행동을 결정하는 정책함수는 로짓기반의 softmax 함수**
- **Policy gradient theorem을 기반으로 정책의 파라미터(가중치)를 업데이트하여 Agent가 더 높은 기대 보상을 얻을 확률이 높은 행동을 선택하도록 정책 개선**
- **수식)**
  - 정책 함수

$$\pi_\theta(a_t \mid s_t) = \frac{\exp(\beta \cdot Q_\theta(s_t, a_t))}{\sum_{a'} \exp(\beta \cdot Q_\theta(s_t, a_t'))}$$

$$Q_\theta(s_t, a_t) = \phi(s_t, a_t)^T \theta$$

- β: 민감도
- Q_θ : 가중치 θ 에 따른 행동 가치 함수
- $\phi(s, a)$: state s 에서 action a 선택 시 해당 대안의 이력정보(features)

  - 정책 개선

$$\theta \leftarrow \theta + \alpha G_t \nabla_\theta \ln \pi_\theta(a_t \mid s_t)$$

$$\nabla_\theta \ln \pi_\theta(a_t \mid s_t) = \phi(s_t, a_t) - \mathbb{E}_{a_t' \sim \pi_\theta}[\phi(s_t, a_t')]$$

- G_t: 기대 리턴값

구분	자료명	자료형태	설명
Input	feature_dim	[Int]	배차 이력 정보 갯수
	learned_weights	[Array] 1Xfeature_dim	(가중치 갱신 시) 이전에 학습된 Agent의 가중치를 적용
	history	[List]	시점 t에서 action, state, reward, action probabilities
Output	action	[Int]	주어진 state에서 Agent가 선택한 대안
	action probabilities	[List]	주어진 state에서 Agent가 action(대안)을 선택할 확률

user

- **이용자 ID에 따른 이력자료('user_history_df', 'features_df')를 비롯하여 학습 환경(env) 및 학습된 에이전트(agent) 인스턴스를 포함하는 딕셔너리 객체**
- **포함 객체)**
  - user_history_df
  - features_df
  - env
  - agent
  - accuracy : 학습 자료 기반 accuracy(설문 자료양이 적어서 train, test set 구분안함, 향후 업데이트 예정)
  - weights

weights.csv

- **이용자 ID에 따른 행동 가치 함수의 feature별 가중치**
- **본 학습의 agent는 행동 가치 함수가 더 큰 대안을 선택한다는 점에서 효용함수 최대화를 기반의 이산선택 로짓모형과 유사함**
- **즉, 행동 가치 함수의 가중치에 따른 대안의 feature의 가중합 점수가 높을 수록 더 좋은 대안이라고 볼 수 있음**
  - 다만, 추정 자료의 한계로 일부 변수에 대하여 양의 가중치가 나타남(향후 리빙랩 수행에 따라 보완 예정)
  - 최종 결과물은 설문자료에 따른 가중치 임으로 가중합을 반영한 시뮬레이션 수행시 실시간 수요 데이터에 응답자 ID 범위의 가상 ID 부여가 필요함

구분	컬럼	설명
이용자 ID	Id	설문지 응답자별 ID [1~500; int]
가중치	access	호출지에서 승차지점까지 도보 접근시간 [분]
	wait	승차지점에서 대기시간 [분] 차량 위치에서 승차지점까지 차량 접근시간 – 호출지에서 승차지점까지 도보 접근시간, 0보다 크거나 같음
	ivt	차내 통행 시간 [분]
	egress	하차지점에서 최종 목적지까지 도보 접근시간 [분]

향후 계획

- **추정 자료의 한계로 일부 변수에 대하여 양의 가중치가 나타남(향후 리빙랩 수행에 따라 보완 예정)**
  - 본 설문 자료는 리빙랩 거주 500명에게 6가지 가상 상황에 대한 답변을 수집한 것으로 학습결과물의 정확도를 올리기 위해서는 배차 이력자료를 획득할 필요가 있음
  - 또한, 본 강화학습 결과물은 누적된 이용경험에 따른 이용자 선호와 수락행태를 모사하기 위한 것으로 의도적으로 상태별 독립성을 보장하지 않음
  - 즉, 누적된 이용 경험속성이 존재하지 않는 설문지 자료에 대하여 본 모형을 적용하는 경우 추정의 한계가 존재함