

# 교육 과정 소개서.

---

스크래치부터 시작하는 강화학습의 모든 것



## 강의정보

|       |   |
|-------|---|
| 강의장   | 온라인 강의   데스크탑, 노트북, 모바일 등   |
| 수강 기간 | 평생 소장   |
| 상세페이지 | <a href="https://fastcampus.co.kr/data_online_reinforce">https://fastcampus.co.kr/data_online_reinforce</a> |
| 강의시간  | 52시간 45분  |
| 문의    | <a href="#">고객센터</a>  |

## 강의특징

|            |   |
|------------|---|
| 나만의<br>속도로 | 낮이나 새벽이나<br><b>내가 원하는 시간대</b> 에 나의 스케줄대로 수강 |
|------------|---|

|               |   |
|---------------|---|
| 원하는 곳<br>어디서나 | 시간을 쪼개 먼 거리를 오가며<br>오프라인 강의장을 찾을 필요 없이 <b>어디서든 수강</b> |
|---------------|---|

|           |  |
|-----------|--|
| 무제한<br>복습 | 무엇이든 반복적으로 학습해야<br>내것이 되기에 이해가 안가는 구간 <b>몇번이고 재생</b> |
|-----------|--|



## 강의목표

- 강화학습 모델을 개선하면서 자주 발생하는 오류 및 디버깅 과정을 담아 문제를 해결해 봅니다.
- 불안정하고 유동성이 큰 현업 환경에서 성능 개선에 도움을 주는 아이디어 학습을 해 봅니다.
- 벤치마크 환경을 하나하나 뜯어 보면서 강화학습 구현을 위한 시뮬레이션 환경을 구축해 봅니다.

## 강의요약

- 지도학습, 비지도학습과 비교하며 강화학습만의 특징과 차이점을 학습합니다.
- 강화학습에서 환경을 묘사하는 개념인 MDP를 학습하고 가치함수, 행동가치함수의 수식적 표현을 이해합니다.
- 환경에 대한 MDP 모델과 Dynamic Programming을 이용하여 Agent를 강화하는 알고리즘을 학습합니다.
- 환경에 대한 MDP 모델없이 Agent를 강화하는 알고리즘을 학습합니다.
- 강화학습의 대표적인 방법론 DQN, Actor-Critic을 공부하고 각 방법론에서 성능, 안정성, 효율이 더 향상된 알고리즘과 개선점을 학습합니다.
- 기존의 데이터 셋, 미리 학습한 사전 정보, 사람의 Feedback을 활용하여 데이터 효율을 증가시키고 나아가 실제 문제 해결에 한 걸음 더 가깝게 학습합니다.



## 강사

권태환

과목

- 스크래치부터 시작하는 강화학습의 모든 것

약력

- 현) 카카오브레인 AI 연구원
- 전) NC Soft 강화학습팀 연구원
- [프로젝트 및 연구 경력]
- 리니지 겨울전쟁 AI 개발 프로젝트 진행
- Starcraft2 환경에서 강화학습 기반 AI 연구 및 개발 진행
- NeurIPS 2022 IGLU Challenge 강화학습 부문 우승
- NeurIPS 2021 NetHack Challenge Neural Agent 부문 준우승

이경재

과목

- 스크래치부터 시작하는 강화학습의 모든 것

약력

- 현) 중앙대학교 RAI LAB 조교수 & 연구책임자
- [프로젝트 및 연구 경력]
- Robot Learning for Object Manipulation with Uncertainty-Aware Autonomous Data Generation and Task Optimization (2021 ~ )
- [SW Star Lab] Robot Learning: Efficient, Safe, and Socially-Acceptable Machine Learning (2019 ~ 2020)
- Intelligent Agent System using Reinforcement Learning (2019 ~ 2020)
- Learning-Based Robotic Grasping (2019)
- Robot Learning from Demonstrations with Mixed Qualities (2017 ~ 2019)
- Human-Level Lifelong Machine Learning (2014 ~ 2017)
- Biomimetic Recognition Technology (2013 ~ 2018)
- Human-Centric Networked Robotics Technology (2013 ~ 2016)



CURRICULUM

01.

Part 1.  
Introduction to  
Reinforcement  
Learning

파트별 수강시간 01:01:08

|                                       |
|---------------------------------------|
| <b>Chapter 1. 강화학습이란 무엇인가?</b>        |
| CH 01-01. 강화학습 소개                     |
| <b>Chapter 2. 강화학습의 특징</b>            |
| CH 02-01. 강화학습의 특징과 지도 및 비지도 학습과의 차이점 |
| <b>Chapter 3. 강화학습의 정형화</b>           |
| CH 03-01. 강화학습의 기본 개념                 |
| CH 03-02. 강화에 대한 정성적 설명               |

CURRICULUM

02.

Markov  
Decision  
Process

파트별 수강시간 02:48:53

|  |
|--|
| <b>Chapter 1. MDP의 정의 및 특성</b>                               |
| CH 01-01. Markov Process란 무엇인가                               |
| CH 01-02. Markov Reward Process란 무엇인가                        |
| CH 01-03. Markov Decision Process란 무엇인가                      |
| CH 01-04. Markov Decision Process란 무엇인가                      |
| <b>Chapter 2. Value Function &amp; Action Value Function</b> |
| CH 02-01. Value Function에 대한 Bellman Equation (1)            |
| CH 02-02. Value Function에 대한 Bellman Equation (2)            |
| CH 02-03. Value Function에 대한 Bellman Equation 예제 (1)         |
| CH 02-04. Value Function에 대한 Bellman Equation 예제 (2)         |
| CH 02-05. Action Value Function에 대한 Bellman Equation (1)     |
| CH 02-06. Action-Value Function에 대한 Bellman Equation 예제      |
| CH 02-07. Bellman Equation의 행렬화 (1)                          |
| CH 02-08. Bellman Equation의 행렬화 (2)                          |
| CH 02-09. Bellman Equation의 행렬화 예제                           |
| CH 02-10. Bellman Optimality Equation (1)                    |
| CH 02-11. Bellman Optimality Equation (2)                    |



CURRICULUM

03.

환경에 대한  
MDP 모델과  
Dynamic  
Programming으  
로 Agent를  
강화하는  
알고리즘

파트별 수강시간 03:58:58

|   |
|---|
| <b>Chapter 1. Policy Evaluation</b>       |
| CH 01-01. ValueFunction 계산                |
| CH 01-02. ActionValueFunction 계산          |
| CH 01-03. ContractionMappingTheorem       |
| CH 01-04. AnalyzationOfPE                 |
| CH 01-05. PolicyEvaluationExample1 (1)    |
| CH 01-06. PolicyEvaluationExample1 (2)    |
| CH 01-07. PolicyEvaluationExample1 (3)    |
| CH 01-08. PolicyEvaluationExample2        |
| <b>Chapter 2. Policy Improvement</b>      |
| CH 02-01. PolicyImprovement               |
| <b>Chapter 3. Policy Iteration</b>        |
| CH 03-01. PolicyIteration                 |
| CH 03-02. PolicyIterationExample1         |
| CH 03-03. PolicyIterationExample2         |
| CH 03-04. PolicyIterationExample3 (1)     |
| CH 03-05. PolicyIterationExample3 (2)     |
| <b>Chapter 4. Value Iteration</b>         |
| CH 04-01. ValueIteration                  |
| CH 04-02. ConvergenceOfValueIteration (1) |
| CH 04-03. ConvergenceOfValueIteration (2) |
| CH 04-04. ValueIterationExamples          |



CURRICULUM

04.

Model-free learning & TD learning

파트별 수강시간 07:20:00

|  |
|--|
| <b>Chapter 1. Monte Carlo Method</b>             |
| CH 01-01. Monte-CarloPrediction (1)              |
| CH 01-02. Monte-CarloPrediction (2)              |
| CH 01-03. Monte-CarloControl                     |
| CH 01-04. AnalysisOfEpsilonSoftPolicyControl (1) |
| CH 01-05. AnalysisOfEpsilonSoftPolicyControl (2) |
| CH 01-06. Monte-CarloMethodExample1 (1)          |
| CH 01-07-Monte-CarloMethodExample1 (2)           |
| CH 01-08. Monte-CarloMethodExample2              |
| CH 01-09. Monte-CarloMethodExample3              |
| <b>Chapter 2. TD Prediction</b>                  |
| CH 02-01. TemporalDifference (1)                 |
| CH 02-02. TemporalDifference (2)                 |
| CH 02-03. TemporalDifference (3)                 |
| CH 02-04. ConvergenceOfTDLearning                |
| CH 02-05. TDLearningExample1                     |
| CH 02-06. TDLearningExample2                     |
| CH 02-07. TDLearningExample3                     |
| <b>Chapter 3. TD Control</b>                     |
| CH 03-01. SARSA                                  |
| CH 03-02. QLearning                              |
| CH 03-03. ConvergenceOfQLearning                 |
| CH 03-04. SARSAExample1                          |
| CH 03-05. SARSAExample2                          |
| CH 03-06. SARSAExample3                          |
| CH 03-07. QLearningExample1                      |
| CH 03-08. QLearningExample2                      |
| CH 03-09. QLearningExample3                      |

## CURRICULUM

## 04.

# Model-free learning & TD learning

파트별 수강시간 07:20:00

|   |
|---|
| <b>Chapter 4. n-step TD Prediction</b>      |
| CH 04-01. n_step_TDPrediction (1)           |
| CH 04-02. n_step_TDPrediction (2)           |
| CH 04-03. ConvergenceOf_n_step_TDPrediction |
| CH 04-04. n_step_TDPredictionExample1 (1)   |
| CH 04-05.n_step_TDPredictionExample1 (2)    |
| CH 04-06. n_step_TDPredictionExample2       |
| CH 04-07. n_step_TDPredictionExample3       |
| <b>Chapter 5. TD (Lambda)</b>               |
| CH 05-01. TDLambda                          |
| CH 05-02. TDLambdaExample1                  |
| CH 05-03. TDLambdaExample2                  |
| CH 05-04. TDLambdaExample3                  |





CURRICULUM

05.

Deep reinforcement learning - DQN

파트별 수강시간 05:56:29

|  |
|--|
| <b>Chapter 1. Neural Networks for RL</b> |
| CH01-01.ValueFunctionApproximation (1)   |
| CH01-02.ValueFunctionApproximation (2)   |
| CH01-03.PolicyExpression                 |
| <b>Chapter 2. DQN Review</b>             |
| CH02-01.Introduction                     |
| CH02-02.Method (1)                       |
| CH02-03.Method (2)                       |
| CH02-04.Results                          |
| CH02-05.DQNExample1 (1)                  |
| CH02-06.DQNExample1 (2)                  |
| CH02-07.DQNExample1 (3)                  |
| CH02-08.DQNExample2                      |
| <b>Chapter 3. Double DQN Review</b>      |
| CH03-01. Introduction                    |
| CH03-02. Background                      |
| CH03-03. Method                          |
| CH03-04. Results (1)                     |
| CH03-05. Results (2)                     |
| CH03-06. DoubleDQNExample1               |
| CH03-07. DoubleDQNExample2               |
| <b>Chapter 4. PER-Review</b>             |
| CH04-01.Introduction                     |
| CH04-02.Method                           |
| CH04-03.Results                          |
| CH04-04.PERExample1 (1)                  |
| CH04-05.PERExample1 (2)                  |
| CH04-06.PERExample2                      |
| Chapter 5. DuelingDQN-Review             |
| <b>CH05-01.Introduction</b>              |
| CH05-02.Method                           |
| CH05-03.Results (1)                      |
| CH05-04.Results (2)                      |
| CH05-05.DuelingDQNExample1               |
| CH05-06.DuelingDQNExample2               |



CURRICULUM

06.

Deep reinforcement learning - PG

파트별 수강시간 14:29:33

|  |
|--|
| <b>Chapter01. Policy Gradient Methods</b>    |
| CH01-01-PolicyGradient (1)                   |
| CH01-02-PolicyGradient (2)                   |
| CH01-03-MonteCarloPolicyGradient             |
| CH01-04-ActorCritic (1)                      |
| CH01-05-ActorCritic (2)                      |
| CH01-06-ActorCritic (3)                      |
| CH01-07-MonteCarloPolicyGradientExample (1)  |
| CH01-08-MonteCarloPolicyGradientExample (2)  |
| CH01-09-ActorCriticExample (1)               |
| CH01-10-ActorCriticExample (2)               |
| CH01-11-ActorCriticExample (3)               |
| CH01-12-AdvantageFunction                    |
| CH01-13-GAE                                  |
| CH01-14. A2CExample (1)                      |
| CH01-15. A2CExample (2)                      |
| CH01-16-GAExample                            |
| <b>Chapter02. Off-Policy Policy Gradient</b> |
| CH02-01-OffPolicyMonteCarloPolicyGradient    |
| CH02-02-OffPolicyActorCritic                 |
| <b>Chapter03. ACER Review</b>                |
| CH03-01-Introduction                         |
| CH03-02-Method (1)                           |
| CH03-03-Method (2)                           |
| CH03-04-Results                              |
| CH03-05-AppendixAnalysisOfRetrace (1)        |
| CH03-06-AppendixAnalysisOfRetrace (2)        |
| CH03-07-AppendixAnalysisOfRetrace (3)        |
| CH03-08-AppendixAnalysisOfRetrace (4)        |
| CH03-09-ACERExample (1)                      |
| CH03-10-ACERExample (2)                      |
| CH03-11-ACERExample (3)                      |

## CURRICULUM

## 06.

# Deep reinforcement learning - PG

파트별 수강시간 14:29:33

|   |
|---|
| <b>Chapter04. PPO Review</b>            |
| CH04-01-Introduction                    |
| CH04-02-Method (1)                      |
| CH04-03-Method (2)                      |
| CH04-04-Results                         |
| CH04-05-PPOExample (1)                  |
| CH04-06-PPOExample (2)                  |
| <b>Chapter05.IMPALA Review</b>          |
| CH05-01-Introduction                    |
| CH05-02-Method (1)                      |
| CH05-03-Method (2)                      |
| CH05-04-Method (3)                      |
| CH05-05-Results (1)                     |
| CH05-06-Results (2)                     |
| CH05-07-Results (3)                     |
| CH05-08-Results (4)                     |
| CH05-09-AppendixConvergenceOfVtrace (1) |
| CH05-10-AppendixConvergenceOfVtrace (2) |
| CH05-11-AppendixConvergenceOfVtrace (3) |
| CH05-12-AppendixConvergenceOfVtrace (4) |
| CH05-13-IMPALAEExample (1)              |
| CH05-14-IMPALAEExample (2)              |
| <b>Chapter06. Distributed Training</b>  |
| CH06-1-Introduction                     |
| CH06-02-DistributedTrainingExample1 (1) |
| CH06-03-DistributedTrainingExample1 (2) |
| CH06-04-DistributedTrainingExample2 (1) |
| CH06-05-DistributedTrainingExample2 (2) |



CURRICULUM

07.

**Advanced Deep Reinforcement Learning**

파트별 수강시간 17:10:22

|  |
|--|
| <b>Chapter01. 직접 해보는 시뮬레이션 환경 구축</b>                 |
| CH01-01.상태, 행동, 보상, 전이확률 정의_이론                       |
| CH01-02.상태, 행동, 보상, 전이확률 정의_실습                       |
| CH01-03.밀바닥 부터 PPO 구현_리뷰                             |
| CH01-04.밀바닥 부터 PPO 구현_실습                             |
| CH01-05.밀바닥 부터 Actor-Critic 구현_이론                    |
| CH01-06.밀바닥 부터 Actor-Critic 구현_실습                    |
| <b>Chapter02. Model Free Offline RL</b>              |
| CH02-01.Conservative Q-Learning - 개념 및 증명            |
| CH02-02.Conservative Q-Learning - 알고리즘 설명            |
| CH02-03.Conservative Q-Learning - 실습                 |
| CH02-04.TD3+Behavior Cloning - 이론                    |
| CH02-05.TD3+Behavior Cloning - 실습                    |
| CH02-06.Offline RL with IQL - 개념 및 증명                |
| CH02-07.Offline RL with IQL - 알고리즘 설명                |
| CH02-08.Offline RL with IQL - 실습                     |
| CH02-09.Mildly Conservative Q-Learning - 개념 및 증명     |
| CH02-10.Mildly Conservative Q-Learning - 알고리즘 설명     |
| CH02-11.Mildly Conservative Q-Learning - 실습          |
| <b>Chapter03. Model Based Offline RL</b>             |
| CH03-01.Model-based Offline Policy Optimization - 이론 |
| CH03-02.Model-based Offline Policy Optimization - 실습 |
| CH03-03.Model Based Offline RL_RAMBO_이론              |
| CH03-04. Model Based Offline RL_RAMBO_실습             |

## CURRICULUM

## 07.

# Advanced Deep Reinforcement Learning

파트별 수강시간 17:10:22

|   |
|---|
| <b>Chapter04. Meta RL</b>                                 |
| CH04-01. Meta RL_MAML_이론                                  |
| CH04-02. Meta RL_MAML_실습                                  |
| CH04-03. Meta RL_MAESN_이론                                 |
| CH04-04. Meta RL_MAESN_실습                                 |
| CH04-05. Meta RL_PEARL_이론                                 |
| CH04-06. Meta RL_PEARL_실습                                 |
| CH04-07. Meta RL_DIAYN_이론                                 |
| CH04-08. Meta RL_DIAYN_실습                                 |
| <b>Chapter05. Preference-Based RL</b>                     |
| CH05-01. Reinforcement Learning from Human Preferences_이론 |
| CH05-02. Reinforcement Learning from Human Preferences_실습 |
| CH05-03. PbRL_PEBBLE_이론                                   |
| CH05-04. PbRL_PEBBLE_실습                                   |
| CH05-05. PbRL_Laerning_from_ranking_이론                    |
| CH05-06. PbRL_Learning_from_ranking_실습                    |
| CH05-07. PbRL_Meta_Reward_Net_이론                          |
| CH05-08. PbRL_Meta_Reward_Net_실습                          |

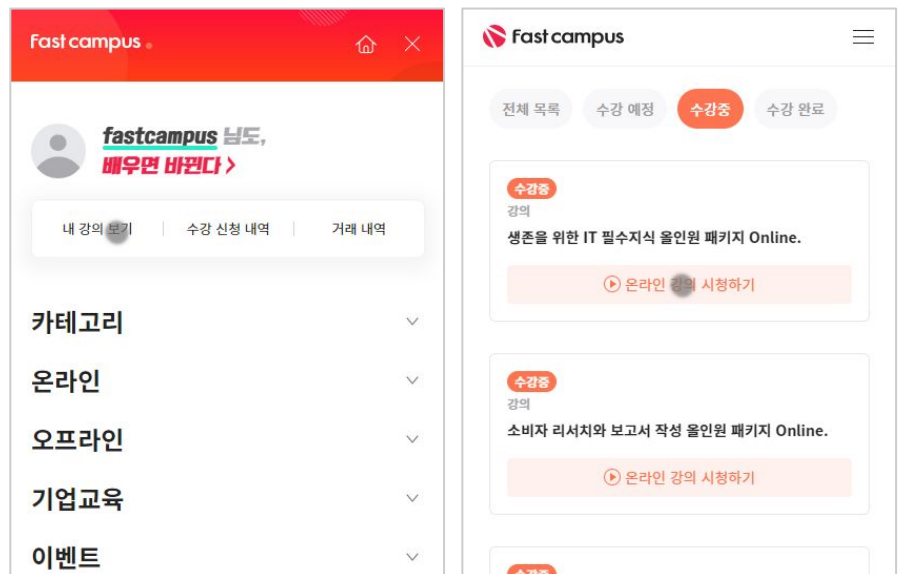


## 주의 사항

- 상황에 따라 사전 공지 없이 할인이 조기 마감되거나 연장될 수 있습니다.
- 패스트캠퍼스의 모든 온라인 강의는 아이디 공유를 금지하고 있으며 1개의 아이디로 여러 명이 수강하실 수 없습니다.
- 별도의 주의사항은 각 강의 상세페이지에서 확인하실 수 있습니다.

## 수강 방법

- 패스트캠퍼스는 크롬 브라우저에 최적화 되어있습니다.
- 사전 예약 판매 중인 강의의 경우 1차 공개일정에 맞춰 '온라인 강의 시청하기'가 활성화됩니다.



## 환불 규정

- 온라인 강의는 각 과정 별 '정상 수강기간(유료수강기간)'과 정상 수강기간 이후의 '복습 수강기간(무료수강기간)'으로 구성됩니다.
- 환불금액은 실제 결제금액을 기준으로 계산됩니다.

|               |                                       |
|---------------|---------------------------------------|
| 수강 시작 후 7일 이내 | 100% 환불 가능 (단, 수강하셨다면 수강 분량만큼 차감)     |
| 수강 시작 후 7일 경과 | 정상(유료) 수강기간 대비 잔여일에 대해 환불규정에 따라 환불 가능 |

※ 강의별 환불규정이 상이할 수 있으므로 각 강의 상세페이지를 확인해 주세요.