
A Survey of Meta-Reinforcement Learning

2023. 04. 03

목차

1. Introduction → 4/3

2. Background → 4/3

- Reinforcement Learning X
- Example algorithms
- Meta-RL Definition
- Problem Categories

3. Few-Shot Meta-RL

- Parameterized Policy Gradient Methods
- Black Box Methods
- Task Inference Methods
- Exploration and Meta-Exploration
- Bayes-Adaptive Optimality
- Supervision

4. Many-Shot Meta-RL

- Multi-Task Many-Shot Meta-RL
- Single-Task Many-Shot Meta-RL
- Many-Shot Meta-RL Methods

5. Application

- Robotics
- Multi-Agent RL

6. Open Problems

- Few-Shot Meta-RL: Generalization to Broader Task Distributions
- Many-Shot Meta-RL: Optimization Issues and Standard Benchmarks

7. Conclusion

1. Introduction

Introduction

❖ A Survey of Meta-Reinforcement Learning (2023, arXiv)

- 옥스포드, 스탠포드 대학에서 작성되었으며, 2023년 arXiv에 게재됨
- 김재훈 연구원과 함께 해당 논문에 관한 내용이 요약된 PPT 작성하면서 관심 연구 도메인 탐색 하려함

A Survey of Meta-Reinforcement Learning

Jacob Beck*

`jacob.beck@cs.ox.ac.uk`
University of Oxford

Risto Vuorio*

`risto.vuorio@cs.ox.ac.uk`
University of Oxford

Evan Zheran Liu

`evanliu@cs.stanford.edu`
Stanford University

Zheng Xiong

`zheng.xiong@cs.ox.ac.uk`
University of Oxford

Luisa Zintgraf†

`zintgraf@deepmind.com`
University of Oxford

Chelsea Finn

`cbfinn@cs.stanford.edu`
Stanford University

Shimon Whiteson

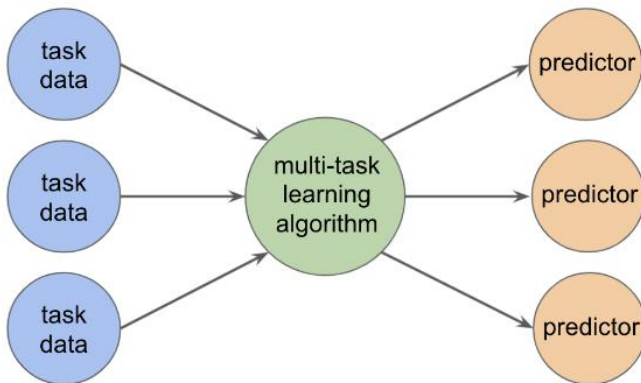
`shimon.whiteson@cs.ox.ac.uk`
University of Oxford

Introduction

❖ Meta Learning

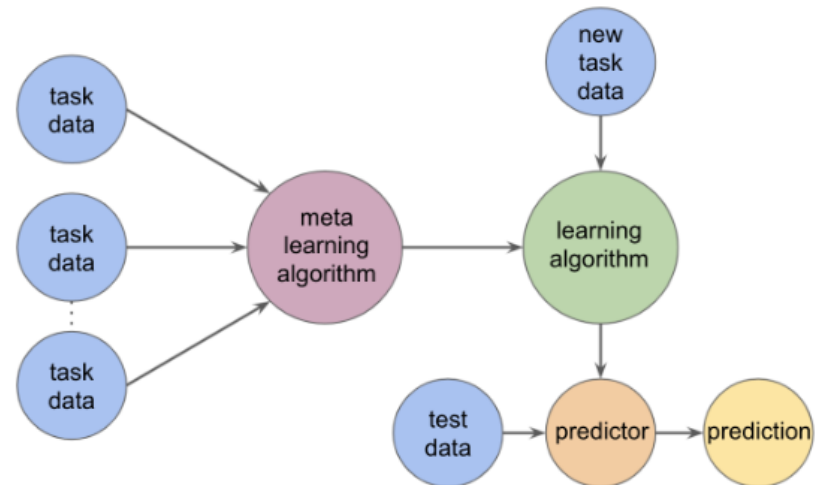
- 다른 task를 위해 학습된 AI 모델을 활용해, 적은 데이터 셋을 가지는 다른 task에도 잘 수행될 수 있도록 학습하는 방식
→ “Learning to learn”
- 최근 각광받고 있는 대부분의 딥 러닝 모델은 많은 양의 데이터가 필요
→ Transfer Learning, Meta-Learning, Multi-task Learning, Continual Learning 등장

Multi-task Learning



각 task에 성능 개선을 위해 동시에 여러 task를 학습하는 것에 초점을 맞춤

Meta-Learning



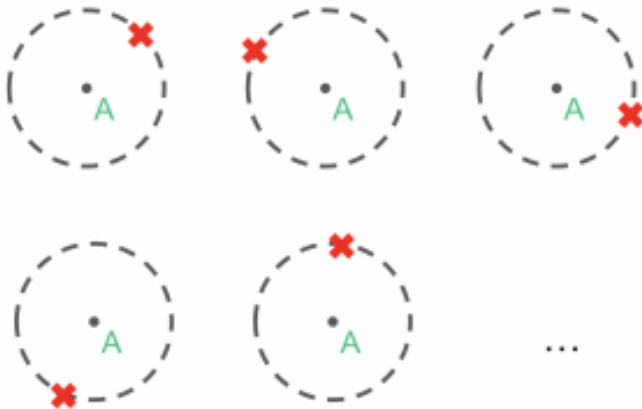
효과적이고 빠르게 학습하는 방법을 학습하는 것에 초점을 맞춤

Introduction

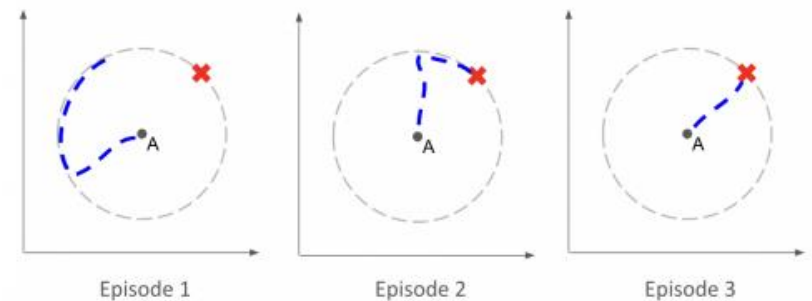
❖ Meta Reinforcement Learning (Meta-RL)

- 학습된 알고리즘이 강화학습 알고리즘이라는 성질을 지닌 meta-learning의 특별한 경우에 해당
- 강화학습이 지닌 한계점들 중 하나인 데이터 효율성 측면 개선 가능 및 다루기 힘든 문제에 관한 최적 해결책 제공 가능

Meta-Training Tasks



Rollout at Meta-Test Time



Learning to learn → Learning to Reinforcement learn

2. Background

Background

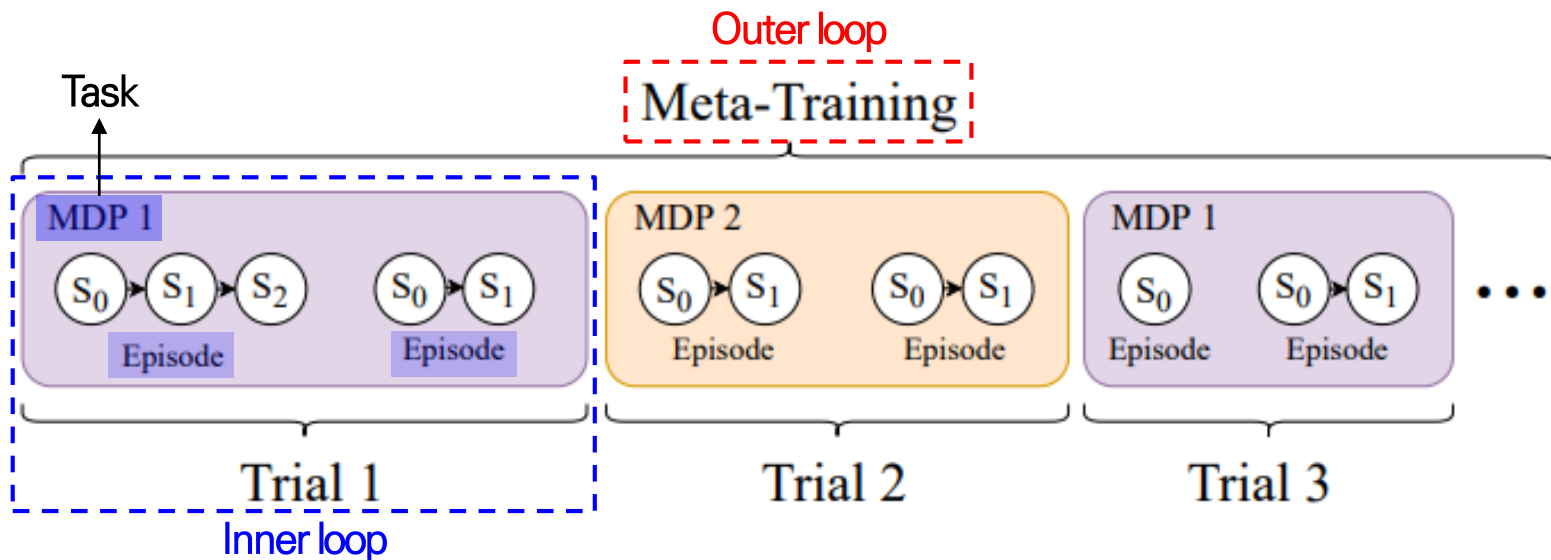
❖ Example Algorithms

- RL은 최적 정책을 학습하고 Meta-RL은 RL을 학습 → bi-level structure (outer loop & inner loop)
 - ✓ Outer loop: Meta-training 과정 / task distribution에서 task를 샘플링
 - ✓ Inner loop: 각 task에 맞는 adaption을 수행하는 과정 / 각 task에 맞는 정책 최적화

Background

❖ Meta-RL Definition

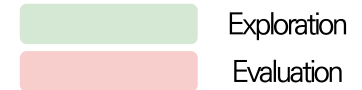
- RL은 최적 정책을 학습하고 Meta-RL은 RL을 학습 → bi-level structure (outer loop & inner loop)
 - ✓ Outer loop: Meta-training 과정 / task distribution에서 task를 샘플링
 - ✓ Inner loop: 각 task에 맞는 adaption을 수행하는 과정 / 각 task에 맞는 정책 최적화



$$\text{Meta-RL Objective: } \mathcal{J}(\theta) = \mathbb{E}_{\mathcal{M}^i \sim p(\mathcal{M})} \left[\mathbb{E}_{\mathcal{D}} \left[\sum_{\tau \in \mathcal{D}_{K:H}} G(\tau) \middle| f_{\theta}, \mathcal{M}^i \right] \right]$$

Background

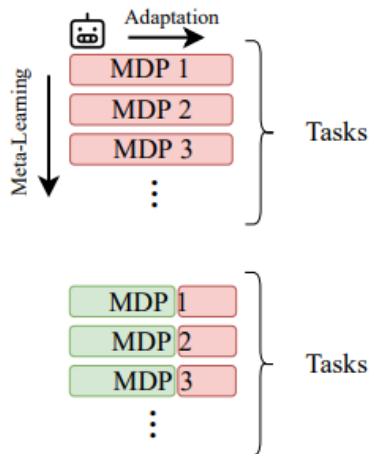
❖ Problem Categories



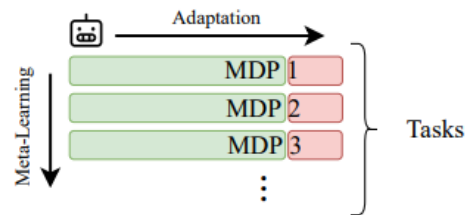
- 에피소드(=샘플)와 task의 수에 따라 3 개의 카테고리 분류

1. Few-shot multi-task: 적은 수의 에피소드를 사용해 새로운 task를 학습하는 것이 목표
2. Many-shot multi-task: 새로운 task를 학습하는데 있어서 표준 RL 알고리즘보다 더 잘 하는 것이 목표
3. Many-shot single-task: 표준 RL 알고리즘 고도화

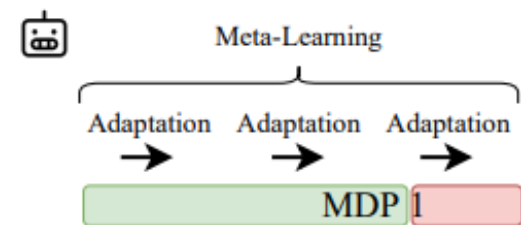
Few-shot multi-task



Many-shot multi-task



Many-shot Single-task



Methods

L2RL(Zero-Shot), VariBAD(Zero-Shot),
MAML(Few-Shot), DREAM(Few-Shot)

LPG, MetaGenRL

STACX, FRODO

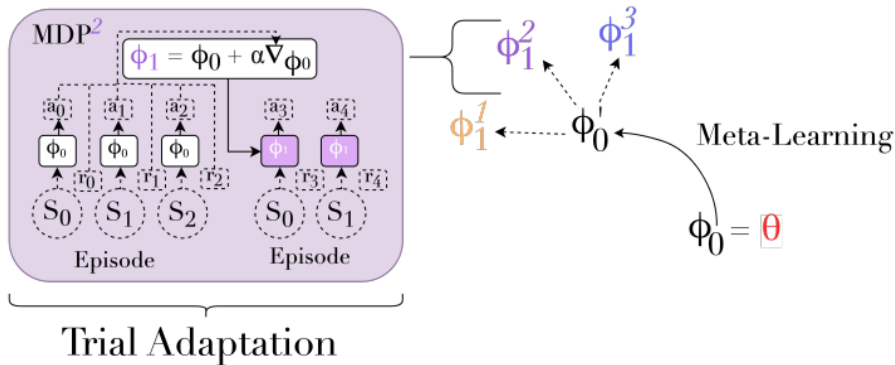
감사합니다.

Background

❖ Example Algorithms

- Meta-RL Objective를 최적화하기 위해 사용되는 표준 meta-RL 알고리즘
 - ✓ *Model Agnostic Meta – Learning (MAML)*
 - ✓ RL^2

Model Agnostic Meta – Learning (MAML)



RL^2

