한국컴퓨터정보학회지
제28권 제1호, 2020. 06.

비지도 강화학습 기반의 인공지능 융합 기술 연구 및 동향 (Research & Trends for Converged AI Technology based on Unsupervised Reinforcement Learning)

• 김민석 (상명대학교 휴먼지능로봇공학과)

I. Introduction

최근 다양한 연구 기술을 상승적 결합하여 새로운 기술 혁신을 추구하는 융합연구가 많은 분야에 걸쳐 진행 중이다. 특히 제4차 산업 혁명 핵심 키워드인 정보통신기술(ICT: Information & Communications Technology) 분야에서는 인공지능(AI: Artificial Intelligence)과 더불어 전 분야의 창의적 융합 포인트를 강조하고 있다. 특히 인공지능의 지능학습인 기계학습(Machine Learning)은 기본적으로 지도학습, 비지도학습 및 강화학습 세 개의 분야로 구분되어 발전하고 있다. 하지만 다양한 방법으로 인공지능을 전 분야에 걸쳐 적용하다 보니 더욱 효과적인 결과를 찾기 위한 새로운 방법들을 강구하기 시작하였고, 이로 인해 기존에 구분되었던 기계학습 분야의 경계면이 점차 허물어져 가는 방향으로 연구 및 개발이 진행 중이다. 예를 들어, 비지도 학습 기반의 클러스터링 기법과 지도학습의 레이블 데이터를 결합한 후 이를 분류 기법에 적용하여 높은 효과를 만드는 준지도 학습(Semi-supervised Learning)이 바로 대표적인 인공지능 융합 기계학습 접근 방법 중 하나이다.

강화학습(Reinforcement Learning)도 새로운 기술 개발을 위해 다른 영역의 비지도 학습 기술과 연계하여 접목하기 위한 시도를 하고 있다. 강화학습은 의사결정이 필요한 시스템 환경에서 임의의 에이전트(Agent)가 환경(Environment) 상태 변화의 제어와 보상 학습을 통해 최적화를 달성하기 위한 기계학습 기법이다. 하지만, 기존 강화학습은 잘 설계된 시스템 혹은 시뮬레이션 환경에서 에이전트 행동 제어에 대한 보상(Reward) 학습을 반복하여 학습 최적화를 성취해야 되기 때문에 보상함수(Reward Function)를 설계하기 위한 연구자의 상당한 노력과 시간이 필요하다. 특히, 최근에 많이 제안되고 있는 딥러닝 기반 심층 강화학습(Deep Reinforcement Learning)은 수많은 양의 학습 데이터를 처리하고 평가해야 하는 요구사항이 존재한다. 또한 부족한 리소스와 복잡한 보상함수 설계에 따른 한계로 인해 실제 시스템에 적용하기까지 쉽지 않다. 따라서 이러한 문제점들을 해결하고 개선하기 위해 다양한 강화학습 방법들이 연구 및 개발되고 있는 추세이다.

II. Preliminaries

1. Unsupervised RL with Intrinsic Reward Function

일반적으로 강화학습은 시뮬레이션이나 시스템 환경으로부터 제공되는 상태 변화에 따른 외적 보상함수(Extrinsic Reward Function)를 이용하여 환경 변화에 대한 에이전트의 행동을 점차 개선하는 방법이다. 하지만, 외적 보상함수에 정해진 환경 상태(States) 값에 학습을 의존하다 보니 에이전트가 복잡한 현상에 따른 행동 제어를 하기에 무리가 따르는 경우가 발생할 수 있다. 따라서 이러한 현상을 극복하기 위해 외적 보상함수 이외에 내적 보상함수(Intrinsic Reward Function)를 사용하는 강화학습 방법이 제안되고 있다.

내적 보상함수는 시스템 환경에서 나타나는 외적 보상함수에서 설계된 학습 요구사항이 아닌 환경에 새롭게 내재 되어 있는 현상을 학습하여 제안하는 방법으로 기존 강화학습에 비해보다 광범위한 작업을 선택하여 학습할 수 있는 지도 강화학습(Unsupervised RL) 방법이다. 또한 에이전트의 행동 개선을 유도하기 위해 시스템 환경 변화에서 나타나는 상호작용(Environmental Interaction)을 강조하고 예측 가능한 환경

변화 요소에 따른 학습 최적화를 성취하는데 목표가 있다.

2. Unsupervised Meta-RL

위에서 언급했듯이 강화학습 알고리즘은 기본적으로 주어진 환경에서 새로운 문제를 빠르게 학습하여 문제를 해결하는 기계학습 방법이다. 따라서 일반적으로 알고리즘의 최적화를 위해서 학습 정책(Policy) 혹은 데이터 학습의 파라미터를 조정하는 하이퍼파라미터튜닝(Hyper Parameters Tuning) 방식을 채택하여 학습 개선을 시도한다. 비지도 메타 강화학습(Unsupervised Meta-RL)은 이러한 방법과 별개로 에이전트가 환경에서 발생하는 환경 변화 요소 혹은 강화학습 속성을 메타(Meta) 테스트/학습하는 방식으로 기존에 학습하지 않는 보상함수나 정책(Policy)을 학습하여 빠르고 효과적으로 성능 최적화하는 방법이다. 일반적으로 강화학습은 학습 최적화를 위해 많은 시간을 초기 학습에 투자해야되며, 에피소드가 끝난 후에도 학습 파라미터를 미세 조정하여 학습을 개선해야 하는 단점이 존재한다. 하지만 메타 학습 방법을 통해 학습 및 테스트를 동시에 진행하여 초기 학습 과정을 생략하고 학습 성능을 크게 향상시킨다. 또한 보상함수 및 정책과 연관된 상호작용을 분리하여 학습 및 처리 과정에서 나타나는 시간적 연속 관계를 이용하여 학습 선순환을 반복하는 방법이다. 메타 테스트 및 학습을 위한 시간은 보상함수와 직접적인 연관성 없지만, 학습 환경의 동적 상호 작용이 있다는 가정하에 환경 개체 정보 없이 학습하는 비지도 학습 방법을 부분적으로 채택하고 있다. 이러한 비지도 메타 강화학습 방법은 해당 시간의 환경 요소를 사용하여 보상 기능으로부터 학습을 최적화를 시킬수 있다는 점에서 기존 강화학습보다 매우 효과적인 학습 방법을 제시되고 있다.

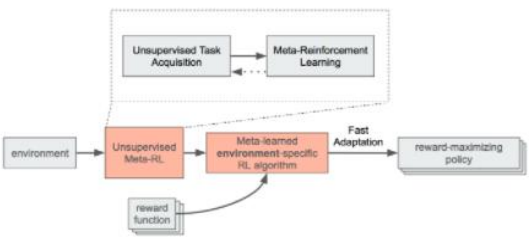

Fig. 1. Unsupervised Meta-RL

III. Convergence of Application Skill

1. Robot Simulation

내적 보상함수를 이용한 비지도 강화학습 방법은 시뮬레이션 환경에서 학습을 진행할 수 있기 때문에 로봇 설계와 같은 복잡한 문제도 다양한 방법의 시뮬레이션을 통해 해결할 수 있다. 구글 로봇팀(Google Brain team and the Robotics at Google team)에서 개발 중인 Dynamics-Aware Unsupervised Discovery of Skills (DADS) 기술에 따르면 비지도 강화학습은 최적화 목표를 위한 최선책 중 하나로 예측 가능성(Predictablilty)의 개념을 강조하고 있다. 예측 가능성은 학습을 통해 환경에서 얻을 수 있는 예측 가능한 변화를 가져오는 것이라고 정의할 수 있다.

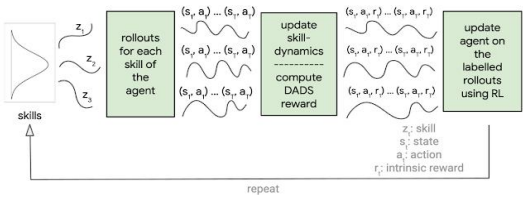

Fig. 2. Workflow for Unsupervised RL(DADS)

DADS는 내적 보상함수를 기반으로 예측 가능한 다양한 기술들을 접목하기 위해 많은 시도를 하고 있다. 내적 보상함수는 환경이 제공하는 변화에 따라 다양한 기술로 전환될 수 있고 환경에 제공되는 기술에 따라 예측 가능한 환경의 변화를 적용하여 학습을 할 수 있다. 환경의 변화에 따른 강화학습 보상함수를 설계하는 이전 방식은 오히려 잠재적 예측 다양성을 억제할 수 있어서, 에이전트가 보다 다양한 잠재적 행동을 포착하고 행동을 시도할 수 있는 내적 보상함수 요소를 추가하여 기술의 확장성이 높이는 기법을 채택하고 있다.

DADS에서는 Skill-dynamics 불리는 신경망을 추가하여 학습한다. 이 신경망은 환경이 제공하는 상태의 동적 변화를 감지하여 학습하는 보조의 신경망이다. 이 신경망을 통해 환경 상태 변화(내적 보상)를 주기적으로 학습하고 예측할 수 있어서 기존 강화학습법보다 좋은 성능을 기대할 수 있다.

Fig. 3. Unsupervised RL using DADS

또한 알고리즘은 다중 에이전트가 활동하면서 얻는 환경에 따른 상호 작용 요소들을 서로 공유하여 학습 속도를 증가시킬 수 있다. 이러한 방법은 휴머로이드 로봇과 같은 고차원의 연속 제어 환경에서도 매우 유용하게 사용할 수 있을 뿐만 아니라 외적 보상함수의 환경 제한을 직접적으로 영향 받지 않기 때문에 다양한 학습 및 예측 제어를 할 수 있는 장점이 있다.

2. Navigation for Path-plannig

우리가 일반적으로 사용하고 있는 지도 경로 탐색 방법에서도 비지도 강화학습을 적용할 수 있다. 지도 경로 탐색에서 가장 중요한 점은 적은 양의 데이터를 이용해서 가장 빠른 경로 탐색(Path Planning)을 제공하는 것이다. 따라서 비지도 메타 강화학습 방법은 에이전트(사용자)의 경로 탐색을 효과적으로 제공하여 목표지점까지 빠르게 도달할 수 있도록 유도하는 학습 가속화 방법을 사용하고 있다.

이 방법은 메타 학습을 통해 경로 탐색을 유도하는 탐색 방법으로 마스터 정책(Master Policy)의 최적화를 위해 하위 정책(Sub-Policy)을 미리 학습하여 정보를 최적화하고 이를 마스터 정책에 전달하여 업데이트하는 경로 탐색을 최적화하는 방식이다. 이때 에이전트는 환경 개체 정보를 제공 받지 않고 오직 경로를 위한 하위 정책 프로세스와 일부 보상만으로 학습을 진행하기 때문에 이는 비지도 강화학습으로 구분할 수 있다. 물론 기존 강화학습 기법 중에도 임의 탐색(Random Exploration)이란 방법을 통해 경로 학습을 임의로 진행하는 학습을 유도하는 방법이 있지만, 하위 정책을 따로 학습하는 메타 학습과는 달리 무작위로 경로를 탐색하여 그 정보를 업데이트하는 방식이기 때문에, 학습 속도가 매우 느리다는 단점이 있다.

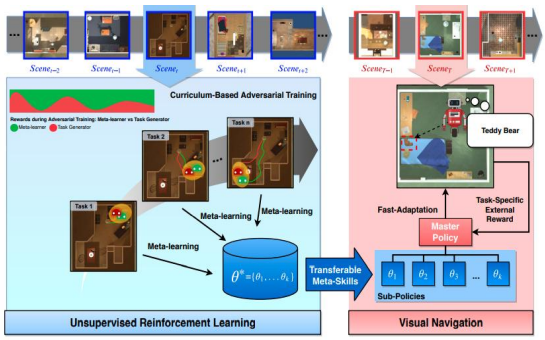

Fig. 4. Overview of Unsupervised Meta-RL

결과적으로 비지도 메타 강화학습은 메타 학습을 통해 얻어지는 보상 값을 기반으로 하위 정책을 학습한 후 주요 경로 탐색을 위한 마스터 정책과 정보를 공유하여 최적화 경로 탐색을 안내한다. 또한 마스터 정책은 하위 정책의 실행 순서를 결정하고 에이전트가 메타 학습 작업에 빠르게 적응할 수 있도록 선순환 과정을 유도하는 학습 방법으로 지도 및 비지도 강화학습을 모두 채택하여 사용하는 매우 효과적인 방법이라고 할 수 있다.

IV. Conclusions

위에서 살펴본 내용과 같이 비지도 강화학습은 다양한 예측 가능성, 보상함수의 확장성, 시간적 효율성, 학습 안정성 등을 제공하기 위한 방법이며 기술 확장을 목표로 연구 및 개발되고 있다. 비록 강화학습은 여전히 학습 보상함수의 정의나 환경 설정의 제약으로 실제 환경이나 시스템에 적용하기 매우 어려운 분야지만, 다양한 방법을 통해 이를 극복하기 위한 지속적인 연구 및 개발이 진행되고 있다. 이러한 연구 개발의 흥미로운 발견은 기술 간의 관계성을 연구하고 학습 제어기술의 구간을 세분화하여 단계별로 융합 발전할 수 있도록 시도하는 부분이며 향후 전반적인 인공지능 기술 분야에 걸쳐 질적 및 양적 기술 향상을 가져올 것으로 기대한다.

REFERENCES

[1] Juncheng Li, Xin Wang, Siliang Tang, etc. "Unsupervised Reinforcement Learning of

Transferable Meta-Skills for Embodied Navigation", Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10.1109/CVPR42600.2020.01214

[2] Abhishek Gupta, Benjamin Eysenbach, Chelsea Finn, and Sergey Levine, "Unsupervised Meta-Learning for Reinforcement Learning", arXiv:1806.04640

[3] Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, etc. "Asynchronous Methods for Deep Reinforcement Learning", Proceedings of The 33rd International Conference on Machine Learning, PMLR 48:1928-1937, 2016

[4] Colin FyfePei Ling Lai, "Reinforcement Learning Reward Functions for Unsupervised Learning", ISNN 2007: Advances in Neural Networks – ISNN 2007 pp 397-402

[5] DADS: Unsupervised Reinforcement Learning for Skill Discovery Project, *https://ai.googleblog.com/2020/05/dads-unsupervised-reinforcement.html*

저 자 소 개

Min Suk Kim received his M.S. in Telecommu-nication and Networks from University of Pittsburgh, USA, in 2008 and 2010. He also received Ph.D. in Electrical and Computer Engineering from University of Massachusetts Lowell, USA, in 2010 and 2016, respectively. Dr. Kim joined Engineer of Electronics and Telecommunications Research Institute(ETRI) from 2016 to 2020. Since 2020, he has been an Assistant Professor with Department of Human Intelligence and Robot Engineering, Sangmyung University, Cheonan, Rep.of Korea. He has published on subject of AI fields such as reinforcement learning, machine learning and edge computing. His research involves machine learning, deep learning and reinforcement learning over a edge computing and centralized cloud computing.