1. Basic Information

Title: Playing Atari with Deep Reinforcement Learning

Authors: Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, Martin Riedmiller

Published in: NIPS Deep Learning Workshop (2013), arXiv preprint

Link: arXiv:1312.5602v1

Terminology:

Q-러닝 : 강화학습의 알고리즘

CNN : 합성 곱 신경망

2. Motivation & Problem

연구 동기: 기존 강화학습에는 사람이 직접 미리 정보를 가공한후, 학습하는 방식 🡪 사람이 직접 인식하고, 행동하는 것처럼 만들 수는 없을까? 에 대한 의문

기존 강화학습 시도중, 학습방법에 문제가 많았음(Correlated Data, Non-stationary Targets)

Q-러닝은 '현재 Q값'을 '타겟 Q값'에 가깝게 업데이트하는 방식으로 학습하는데, Q값의 업데이트 과정에서 가중치가 변해 타겟 Q값까지 같이 변함

Correlated Data 🡪 비슷한 프레임이 많아, 특정상황에 과적합됨

문제 정의: 기존 강화학습 방법 ; 사람이 직접 미리 정보를 가공한후, 학습하는 방식의 한계.

3. Methodology

제안된 방법론: Deep Q-Network (DQN), Experience Replay, Fixed Q-Targets

Experience Replay: 모든 경험 (상태, 행동, 보상, 다음 상태)을 Replay memory라는 큰 버퍼에 저장, 이후 무작위로 사용하여 데이터 효율을 높임, 비슷한 데이터가 많더라도, 편향된 학습을 방지함

Fixed Q-Targets: 타겟을 얼려서, 가중치가 바뀌어도, 타겟값이 안바뀌게 함

사용된 데이터셋: 아타리 2600 게임

4. Key Results

기존의 모든 강화학습 알고리즘을 능가하는 성능

Breakout, Enduro, Pong에서는 사람의 점수를 넘음

5. Strengths

Raw Pixel 사용 ( Hand crafted feature 필요없음)

7. Personal Insights

내가 이해한 핵심 아이디어 : Q tareget 값의 freeze로 인한, 기존의 문제점 극복, Experience Replay를 이용한, 기존 과적합 문제 해결