

상태 표현 방식에 따른 심층 강화 학습 기반 캐릭터 제어기의 학습 성능 비교

손채준 , *이윤상

Dept. of Computer Science, Hanyang University
thscowns@gmail.com, *yoonsanglee@hanyang.ac.kr

요 약

물리 시뮬레이션 기반의 캐릭터 동작 제어 문제를 강화학습을 이용하여 해결해 나가는 연구들이 계속해서 진행되고 있다. 이에 따라 이 문제를 강화학습을 이용하여 풀 때, 영향을 미치는 요소에 대한 연구도 계속해서 진행되고 있다. 우리는 지금까지 이뤄지지 않았던 상태 표현 방식에 따른 강화학습에 미치는 영향을 분석하였다. 첫째로, root attached frame, root aligned frame, projected aligned frame 3 가지 좌표계를 정의하였고, 이에 대해 표현된 상태를 이용하여 강화학습에 미치는 영향을 분석하였다. 둘째로, 동역학적 상태를 나타내는 캐릭터 관절의 위치, 각도에 따라 학습에 어떠한 영향을 미치는지 분석하였다.

1. 서론

물리 시뮬레이션 기반의 사람 캐릭터 동작 제어는 상당히 어려운 문제로 여겨져 왔다. 많은 것들이 이 문제를 어렵게 만들고 있지만, 가장 큰 이유는 사람 캐릭터가 많은 관절을 가지고 있어 높은 자유도(degrees of freedom, DOFs)를 가지기 때문에 상태(state)의 차원이 크기 때문이다. 최근에는 딥뉴럴네트워크 정책을 기반으로 하는 강화학습의 발전으로 이 문제를 해결해 나가고 있다. 강화학습을 사용하여 문제를 풀기 위해서는 네트워크 구조, 하이퍼파라미터 튜닝, 상태, 행동(action), 보상(reward)이 문제에 맞게 적절히 설정이 되어야 사람 캐릭터가 걸거나 뛰는 동작을 만들어 낼 수 있다

상태의 경우, 같은 속성을 상태로 정의하더라도 어떤 좌표계에 대해 표현된 상태 값인지에 따라 그 특성이 달라질 수 있다. 그러나 기존 연구들은 상태를 표현하기 위한 좌표계의 선택에 집중하여 분석하지는 않았다. 이 연구에서는 다양한 좌표계에서 표현된 상태가 강화학습을 통한 문제 해결에 어떠한 영향을 미치는지 알아볼 것이다. 또한, 캐릭터의 동작을 나타내는 동역학적 상태를 나타내기 위해 관절 위치 혹은 관절 각도가 사용될 수 있다. 우리는 참조동작을 따라하도록 하는 캐릭터 제어문제를 강화학습을 이용하여 풀 때, 동역학적 상태를 정의하는 방법이 미치는 영향을 분석할 것이다. 이를 통해 우리는 물리 시뮬레이션 기반의 사람 캐릭터 동작 제어 문제를 강화학습을 이용하여 풀 때, 동역학적 상태를 정의하는 가장 좋은 방법 및 상태를 표현하는 좌표계의 선택에 중요한 요소가 무엇인지 보이고자 한다.

2. 방법

우리는 물리 시뮬레이션 기반의 캐릭터를 참조동작을 따라하도록 제어하는 문제를 강화학습을 이용하여 풀 때, 상태를 여러가지 다른 방식으로 정의하고 이것이 학습에 어떠한 영향을 미치는지 확인하였다. 이를 위해 우리는 DeepMimic[1]에서 제안한 강화학습 기반 캐릭터 제어 프레임워크의 구조를 사용하였다. 우리는 DeepMimic[1]에서 정의된 것처럼 관절 위치, 관절 각도, 속도, 각속도, 위상, root 의 높이를 상태로 사용하였다. 첫째로, 우리는 상태가 서로 다른 좌표계에 의해 표현되었을 경우에 학습 성능을 비교하였다. 둘째로, 동역학적 상태가 다르게 정의되었을 경우에 학습 성능을 비교하였다.

2.1 상태를 나타내는 좌표계

어떤 좌표계에 대해 표현된 상태 값인지에 따라 특성이 다르기 때 문에 우리는 총 3 가지의 좌표계를 이용하여 상태를 다르게 표현하여 사용하였다. (1) Root attached frame, (2) Root aligned frame, (3) Projected root aligned frame 세가지 좌표계를 이용하였다. Root attached frame 은 root 의 로컬 좌표계를 의미한다. root aligned frame 은 상체가 바라보는 방향과 y 축 및 이 두가지 축을 외적인것을 각 축으로 가지는 좌표계를 의미하고 위치는 root 의 위치로 정의한다. Projected root aligned frame 은 root aligned frame 의 원점을 xz 평면에 projection 시킨 좌표계이다. 우리는 앞서 설명한 총 3 가지의 좌표계를 이용하여 관절 위치, 관절 각도, 속도, 각속도를 각각의 좌표계에 대해서 표현하여 상태를 정의하였다.

2.2 동역학적 상태 정의

이전 연구에서 동역학적 상태를 정의할 때 관절 위치, 관절 각도를 모두 사용한 연구도 존재하였고, 둘 중 하나만 사용한 연구도 존재했다. 이에 성능을 비교하는 연구도 존재하였으나 관절 위치, 각도 모두 사용한 것과 관절 각도를 사용한 것의 비교였고, 참조 동작을 추적하는 문제가 아닌 사람이 걷는 문제를 강화학습을 이용해 학습하는 경우 성능을 비교하였다. 우리는 관절 위치만 사용한 것, 관절 각도만 사용한 것, 관절 위치, 각도 모두 사용한 것 세가지로 나누어서 참조 동작을 추적하는 문제의 강화학습을 진행하였다. 이 실험에 대해서 모든 상태는 root aligned frame 에 의해 표현된다.

3. 결과

시뮬레이션은 bullet physics 를 이용하였다. 기본적으로 [1]에서 사용된 것처럼 aligned frame 에 대해 표현된 상태를 사용하였고, PPO 를 이용하여 학습을 진행하였다. 하이퍼파라미터는 DeepMimic[1]에서 사용한 값을 그대로 사용하였고, 각 환경마다 160000 iteration(65,000,000 sample) 씩 학습을 진행하였다. 그래프는 저장된 네트워크를 load 하여 50 번씩 에피소드를 진행하여 누적보상의 평균과 표준편차를 계산하였고 이를 이용하여 그래프를 만들었다.

3.1 상태를 나타내는 좌표계에 따른 성능 비교

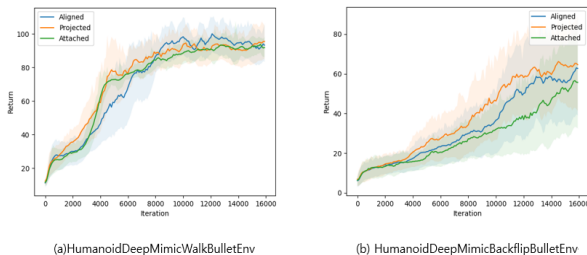


그림 2. 3가지 좌표계에 의해 표현된 상태를 이용하여 학습을 진행하였을 때, 얻어낸 학습 그래프이다.

WalkEnv, BackflipEnv 에 대해 상태를 앞서 언급한 세가지 좌표계에 대해 표현하여 학습을 진행하였다. 그림 2 에서 알 수 있듯이 WalkEnv 에 대해서는 상태를 각 프레임에 따라 표현하는 것이 크게 의미가 있지는 않았다. 그러나, BackflipEnv 에서는 상태를 각각 다른 좌표계에 대해서 표현하는 것이 큰 영향을 미치는 것을 확인할 수 있었다. 학습한 환경들의 참조동작을 살펴보면 WalkEnv 의 경우 root 의 높이는 거의 변하지 않고, 회전이 큰 동작이 없이 앞으로 나아가기만 한다. 반면 BackflipEnv 의 경우 root 의 높이가 계속해서 변하고, 회전이 큰 동작이다. 이러한 환경의 특성 차이로 인해 그림 2(a)와 2(b)에서 나타난 경향성에 차이가 난다.

그림 2(b)를 보면 BackflipEnv 를 학습시킬 때 projected

frame 에 의해 표현된 상태를 이용하여 학습시킨 결과가 다른 좌표계를 이용하였을 경우보다 더 좋은 성능을 보여준다. projected frame 에 의해 표현되는 관절의 위치 정보에는 나머지 좌표계에 의해 표현되는 상태에 비해 root 의 높이가 더해져서 상태로 정의된다. 그렇기에 root 의 높이 정보가 계속해서 달라지는 BackflipEnv 의 경우 projected frame 에 의해 표현된 상태가 학습 속도 큰 영향을 준다.

3.2 동역학적 상태 정의에 따른 성능 비교

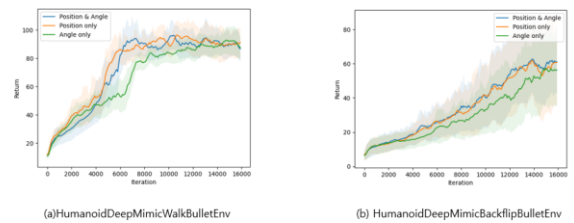


그림 3. 캐릭터의 동역학적 상태를 관절 위치만 사용한 경우, 관절 각도만 사용한 경우, 관절 위치, 각도 모두 사용하여 상태를 정의하고 학습시켜 얻어낸 학습 그래프이다.

결과를 살펴보면 두 환경에서 모두 최종적으로는 비슷한 수준의 누적보상을 보여주었다. 두 환경 모두 위치 정보를 사용하여 학습한 경우가 각도만 사용한 경우보다 빠르게 최종수준에 도달하였다. 이를 통해 환경의 참조 동작이 갖는 특성, 즉 동작의 회전과 root 높이 변화에 상관없이 캐릭터 관절의 위치 정보가 학습에 더 큰 영향을 주는 것을 알 수 있었다. 강화학습에서 에이전트가 캐릭터의 상태를 학습하여 가장 높은 누적보상을 갖도록 행동을 만들어내는데, 에이전트가 캐릭터의 동역학적 상태를 학습할 때, 캐릭터의 위치정보가 에이전트가 학습을 더 쉽게 할 수 있도록 만드는 것으로 여겨진다.

4. 결론

물리 시뮬레이션 기반의 캐릭터 컨트롤 문제를 강화학습을 이용하여 푸는 연구들은 계속해서 이루어져 왔다. 최근에는 참조동작을 따라하도록 환경을 구성하고 이를 강화학습을 이용하여 푸는 연구들이 주를 이뤘다. 이에 우리는 이러한 문제 정의속에서 상태를 표현하는 방식이 미치는 영향에 대해서 알아보았다. 지금까지 알아본 바로는 참조동작이 가지는 특성에 따라서 상태를 정의하는 좌표계를 선택해야한다. 또한, 상태를 정의함에 관절의 위치 정보와 각도가 학습하는데 미치는 영향을 알아보았다. 이 연구를 통해 상태표현방식이 강화학습에 미치는 영향을 알아볼 수 있었다.

참고문헌

[1] X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne, "Deepmimic," ACM Transactions on Graphics, vol. 37, no. 4, p. 1– 14, Aug 2018. [Online]. Available: <http://dx.doi.org/10.1145/3197517.3201311>