

# 인체 근골격 모델의 현실적인 보행 동작 생성을 위한 평균 상태 손실 함수를 이용하는 모델 기반 강화 학습\*

김민관<sup>0</sup>, 이윤상  
한양대학교 컴퓨터소프트웨어학과  
{palkan21, yoonsanglee}@hanyang.ac.kr

## Generating Realistic Locomotion of Musculoskeletal Humanoid with Model Based Reinforcement Learning using Average State Loss Function

Minkwan Kim<sup>0</sup>, Yoonsang Lee  
Department of Computer Software, Hanyang University

### 요약

본 연구는 최근 캐릭터 컨트롤러 학습 연구의 경향성과 다르게, 레퍼런스 모션 데이터의 도움 없이 근골격 모델의 자연스러운 보행 동작을 학습하는 방법을 제안한다. 이를 위해 사람의 구동을 모델링한 근육 시뮬레이션, 보행의 특성을 표현하기에 적합한 손실 함수, 그리고 시뮬레이션을 예측하는 월드 모델 등을 사용했다.

### 1. 서론

최근 심층 강화 학습으로 물리 시뮬레이션 되는 캐릭터가 모션 캡처 데이터를 재현하는 정책을 찾는 연구가 많이 발표되고 있다. 특히, 데이터셋에 포함된 움직임을 낮은 차원의 잠재 공간에 매핑하고, 이를 이용해서 특정 스킬을 선택하고, 움직임의 방향 및 속도 등을 조종하며 더 어려운 작업을 수행하는 능력을 여러 레벨에 걸쳐 학습하는 연구 [1]도 발표된 바 있다.

이러한 데이터 기반의 접근 방식은 다양한 동작 생성에 좋은 모습을 보이나, 학습 도중에 사용한 데이터의 특성이 결과 동작에 내재된다는 한계점이 존재한다. 또한, 모션 캡처가 어려운 동물이나 데이터가 희소한 동작, 혹은 가상의 생물 등에는 적용이 불가능하다.

모션 데이터를 사용하지 않고 동작 제어 정책을 학습하려는 시도도 있었으나, 보상 함수나 각 관절의 토크 제한 등에 대한 세밀한 조절을 필요로 하거나 [2], 모든 각도에서 봤을 때 자연스러운 동작을 만들지 못하고 오랫동안 보행을 지속하지 못하거나 [3], 학습된 정책

이 선형 피드백 규칙과 함께 사용되어야 한다는 [4] 한계점이 있었다. 최근 6년간 모션 데이터에 기반하지 않는 다른 관련 연구를 찾기 힘들 정도로, 데이터 기반 방식에 비해서 동작 생성 정책의 학습에 어려움을 갖는다.

이러한 학습의 어려움을 극복하기 위해, 우리는 시뮬레이션을 예측하는 월드 모델 기반의 엔드 투 엔드 강화 학습으로 시뮬레이션의 내재된 원리를 제어 정책에 전달했다. 또한, 보행 내의 특정 시간 구간 동안 캐릭터의 평균 상태를 계산해서 이를 바탕으로 손실 함수를 계산했으며, 실제 생물이 보행에서 최소화한다고 알려진 근육에 의해 계산된 에너지 소모 수식을 포함한다.

위와 같은 요소들에 힘 입어 본 연구는 데이터의 도움 없이 실제 사람과 비슷한 보행 동작을 생성했으며, 이는 근육 기반의 모델이 존재하는 모든 캐릭터의 보행 동작 생성에 확장될 수 있는 잠재력을 가진다.

### 2. 시스템 요약

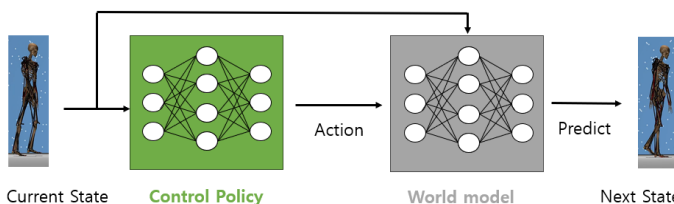


Figure 1: Overview of the proposed system

본 연구는 레퍼런스 모션의 도움 없이 현실적인 사람의 보행 동작을 생성하는 방법을 제안하며, 그림 1은 전체 학습 과정을 나타낸다. 제어 정책은 캐릭터의 현재 상태를 입력 받아서 각 근육의 활성화 값을 출력하고, 월드 모델을 거쳐서 캐릭터의 다음 상태가 결정된다.

시뮬레이션 단계가 미분이 불가능한 문제를 해결하기 위해 모델 기반의 강화 학습을 사용하였으며, 월드 모델은 학습 도중에 물리 엔진의 근육 시뮬레이션 결과로

\* 구두 발표논문, 요약논문 (Extended Abstract)  
\* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회 (논문지)에 제출 준비중임.  
\* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2023-00222776) 및 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광 연구개발사업 지원을 (RS-2024-00399136) 받아 수행되었음.

예측하고 기울기를 역 전파하는 역할을 하고, 런 타입에는 물리 엔진으로 근육 시뮬레이션을 진행했다.

또한, 매 스텝마다 캐릭터의 상태를 평가하는 대신, 특정 시간 구간 동안의 평균 상태를 기반으로 보행에 대한 손실 함수를 계산하였다. 이는 실제 사람의 보행 주기에서 각 관절의 회전이나 루트의 움직임이 특정 값을 중심으로 편차를 갖는 것에서 기반했으며, 실제로 자연스러운 움직임을 학습하는 것에 중요한 역할을 했다.

### 3. 학습 과정 및 시뮬레이션 상세

제안하는 시스템은 우선 시뮬레이션 샘플을 수집하고, 해당 샘플로 월드 모델 및 제어 정책을 번갈아 가며 학습하는 과정을 반복한다.

#### 3.1. 근육 시뮬레이션 샘플 수집

해당 단계에서는 현 시점의 제어 정책이 출력한 근육 활성화 값으로 물리 시뮬레이션을 진행하며 캐릭터의 이전 상태 및 다음 상태를 샘플로 버퍼에 저장한다.

제어 정책이 출력하는 0에서 1사이의 근육 활성화 값은, 각 근육의 이완 및 수축의 정도를 표현한다. 근육의 힘은 근육의 길이, 속도 및 활성화 값 기반으로 미리 정의된 그래프로부터 계산되며, 그 후 각 근육 별로 정의된 경로 점 사이에서 당기는 방향으로 힘이 적용해서 최종적으로 관절을 회전시킨다.

근육 시뮬레이션에는 MuJoCo [5]가 사용되었고, 각 근육이 낼 수 있는 힘, 경로 점의 위치 등 근육 관련 파라미터가 포함된 근골격 모델은 기존 연구 [6]의 것을 사용하였다.

#### 3.2. 시뮬레이션을 예측하는 월드 모델 학습

월드 모델은 현재 캐릭터의 상태와, 제어 정책이 출력한 근육 활성화 값을 입력 받아서 다음 캐릭터의 상태를 출력한다. 월드 모델이 예측한 다음 상태는 버퍼에 저장된 실제 시뮬레이션 결과와 오차가 적어지도록 학습된다. 캐릭터의 상태는 각 근육의 길이, 속도와 각 관절의 위치, 속도, 회전, 각 속도 등의 정보로 구성된다.

#### 3.3. 보행 동작 생성을 위한 제어 정책 학습

제어 정책은 캐릭터가 자연스러운 동작으로 앞으로 넘어지지 않고 이동하는 것을 목표로 학습된다. 정책은 액션을 출력하고 앞서 학습된 월드 모델로 캐릭터 상태를 업데이트하는 것을 24 스텝동안 반복한다. 그 뒤, 이 시간 구간 동안의 평균 상태를 이용해서 보행 동작의 만족 정도를 평가하며, 손실 함수의 구성은 아래와 같다.

$$loss = l_{vel} + l_{height} + l_{dir} + l_{pose} + l_{energy}$$

$l_{vel}$ ,  $l_{height}$ ,  $l_{dir}$ ,  $l_{pose}$ 는 각각 시뮬레이션 구간 동안 루트의 속도, 높이, 회전 및 캐릭터의 자세의 평균 값과 목표 값 사이의 차이를 의미하며, 이 차이를 줄이는 방향으로 학습된다. 이는 실제 사람이 보행 시 속도, 높이, 각 관절의 회전 등이 편차를 보이지만, 보행의 한 주기에 해당하는 시간 구간의 평균을 구하면 특정 값으로

수렴한다는 관찰에서 기반한 손실 함수 계산 방식이다.

실험을 통해 찾은 목표 속도는 일반적인 보행의 평균 속도 범위에 포함되는 1.2m/s, 목표 높이는 캐릭터의 초기 높이보다 약간 낮은 0.91m, 목표 자세는 가만히 서 있는 자세로 일반적인 예측과 비슷하다.  $l_{energy}$ 는 목표 값이 0으로, 근육의 에너지 소모 수식 계산 결과 값을 최소화하며 자연스럽게 안정적인 동작을 학습하는데 중요한 역할을 한다.

### 4. 실험 결과 및 결론

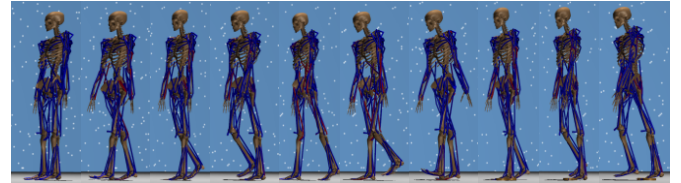


Figure 2: Realistic locomotion from proposed system

우리는 앞서 언급한 3가지 단계를 거쳐서 월드 모델과 제어 정책을 번갈아 가며 학습했고, 최종적으로 그림 2와 같이 실제 사람의 보행에서 보일법한 모습을 갖춘 현실적인 동작을 근육 시뮬레이션으로 생성했다. 제어 정책과 월드 모델은 모두 완전 연결 신경망이며 학습은 NVIDIA 4070 GPU에서 약 15시간이 소요되었다.

또한 섹션 2에서 제안한 요소들 대신 단순히 근육 활성화 값을 최소화하거나 혹은 매 스텝마다 손실 함수로 보행 정도를 평가하도록 학습할 경우, 무릎을 굽힌 채로 걷는 부자연스러운 동작이 생성되는 것을 확인했다.

결론적으로, 우리의 연구는 모델 기반의 강화 학습과 평균 상태 기반의 손실 함수, 그리고 근육에 의해 계산되는 에너지 최소화 요소를 포함하는 구조로 레퍼런스 모션의 도움 없이 자연스러운 보행 동작을 생성할 수 있음을 보였다. 추후에는 다양한 속도의 동작 생성 혹은 사람 이외의 근골격 모델에 대한 확장을 목표로 한다.

### 참고문헌

- [1] Y Feng, X Xu, L Liu, MuscleVAE: Model-Based Controllers of Muscle-Actuated Characters, SA '23: SIGGRAPH Asia 2023 Conference Papers, December 2023, Article No.: 3, Pages 1–11
- [2] W.Yu, G.Turk, K.Liu, Learning Symmetric and Low-energy Locomotion, ACM Transactions on Graphics 37(4), 2018.
- [3] M.K.Kim and Y.S.Lee. Learning Human-like Locomotion Based on Biological Actuation and Rewards, ACM SIGGRAPH 2023 Posters July 2023 Article No.: 5 Pages 1–2
- [4] G.C.Kang and Y.S.Lee. “Finite State Machine-Based Motion-Free Learning of Biped Walking”. IEEE Access 9 (2021), 20662–72.
- [5] Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012.
- [6] Y.S.Lee, M.S.Park, T.S.Kwon, J.H.Lee, Locomotion control for many-muscle humanoid, ACM Transactions on Graphics, 33(6), 2014.