'96 대한전기학회 하계학술대회 논문집 (1996. 7. 22~24)

Reinforcement 학습을 이용한 두발 로보트의 보행 자세 교정

이 건 영
광운 대학교 전기공학과

Gait synthesis of a biped robot using reinforcement learning

Keon Young Yi
Dept. of Electrical Eng., Kwangwoon University

Abstract

A neural network(NN) mechanism is proposed to modify the gait of a biped robot that walks on sloping surface using sensory inputs. The robot starts walking on a surface with no priori knowledge of the inclination of the surface. By accumulating experience during walking, the robot improves its walking gait and finally forms a gait that is adapted to the surface inclination. A neural controller is proposed to control the gait which has 72 reciprocally inhibited and excited neurons. PI control is used for position control, and the neurons are trained by a reinforcement learning mechanism. Experiments of static gait learning and pseudo dynamic learning are performed to show the validity of the proposed reinforcement learning mechanism.

1. 서론

보행형 로보트는 바퀴를 장착한 이동식 로보트에 비하여 표면이 고르지 못한 바닥 조건(계단 포함)에서 우수한 특성을 발휘한다. 특히 다리의 길이를 적절히 조절하여 바닥 형태와 무관하게 물체를 부드럽게 운반하는 능동 완충 능력(active suspension)을 갖는다. 이러한 장점으로 인하여 보행형 로보트는 핵발전소나 해저 탐사 등 극한적인 상황에서 인간을 대신하기에 매우 적합하다. 두발 보행형 로보트의 경우는 바닥 접촉면이 좁아서 좁은 경로를 따라 이동이 가능하여 그 유용성이 배가된다. 이러한 두발 보행형 로보트의 장점을 현실에 적용하기 위하여 많은 연구가 진행되어 왔으나[1], 복잡한 바닥 조건에서도 안정된 보행을 유지해야 하는 기술적 어려움으로 인하여 그 성과는 만족스럽지 못하다.

최근에는 어려운 바닥 조건에서의 보행, 로보트 동력학을 고려한 보행에 관한 연구가 활발히 진행되고 있다. [2],[3]. 본 연구에 앞서 진행된 연구[3]에서는 스위치 센서를 이용한 바닥 면의 경사도 측정으로 두발 보행 로보트의 자세 교정을 하였으나, 이 방법이 측정 잡음에 강인함에도 불구하고 경사면이 자주 변하는 경우(일반적인 경우)에 적용하기 어려운 단점이 있으며, 로보트에 대한 구체적 기구 구조 값이 없으면 자세 교정을 위한 Data 산출이 불가능한 단점이 있다.

본 연구에서는 연구 [3]의 문제를 극복하기 위한 방법의 일환으로 보행 로보트 제어를 위한 신경망을 제시하는데, 이는 기존의 신경망 기법을 실제 응용이 가능하도록 수정하였다. 즉 각 신경들은 주어진 시스템으로부터의 직접 지시나 응답이 아닌 비통제 강화학습(unsupervised reinforcement learning)을 통하여 시스템의 특성을 배우게 된다. 좀더 구체적으로 말하면, 경사면에서의 안정된 자세를 얻기 위하여 어느 관절을 조정해야 효과적인지, 어느 방향으로 조정해야 하는지를 알 수가 없어 적절한 자세 교정을 위한 방법 구현이 어렵다. 물론 목표 보행 조건이 정해지면 무수히 많은 실험을 통하여 보행 자세 교정을 위한 요소를 밝혀 낼 수 있으나, 실제 상황에서는 보행 면이 항상 일정하지 않으며 측정 잡음 등으로 인하여 이 또한 불가능하다. 측정 잡음에 대한 일부 특성이 파악되면 확률 제어 방식으로 보행 자세 교정이 가능하나, 구성될 제어기의 복잡성으로 인하여 실시간 구현이 어렵다. 따라서, 본 연구에서도 시스템 특성을 온라인으로 배우고, 잡음을 학습기에서 감쇄 시키며, 직접적인 피드백 신호를 요하지 않는 [4] 강화학습(reinforcement learning)법을 로보트의 경사면 보행을 위한 자세 교정용 제어기로 선택하였다.

2. SD-2 보행 로보트

제어 대상인 SD-2 보행 로보트는 Zheng에 의하여 미국 Clemson 대학에서 개발되어 현재는 Ohio 주립 대학에 소장되어 있다. 이 절에서는 제어 대상 로보트의 구조와 보행 방법에 대하여 설명한다.

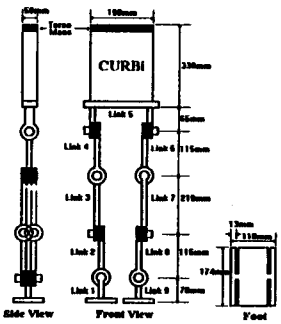

그림 1. SD-2 로보트의 구성도

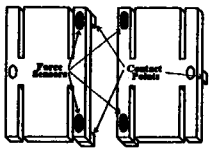

그림 2. 로보트 발과 힘센서

2.1 보행 로보트의 구조

SD-2 로보트는 그림과 같이(그림 1) 8개의 관절과 9개의 연결부로 구성되었다. 4개의 관절은 측면도 상의 운동(전·후 보행), 그리고 나머지 4 관절은 정면도 상의 운동(좌·우 중심 이동)을 담당한다. 각 다리의 위쪽 두 관절과 아래 두 관절은 각각 인간의 엉덩이 관절 및 발목 관절의 기능을 모사 한다. 유의할 점은 이 로보트에는 인간의 무릎 관절을 위한 기능이 없는 것이다.

그림 2는 로보트의 발바닥과 지면과의 접촉부(좌·우 3곳 씩) 및 힘센서(좌·우 2곳씩)를 나타낸다. 강화학습 신호를 위한 힘센서는 Measurements Group Inc.의 Strain gauge로, TK-13-T092P-10C, 발 앞 및 뒤쪽이 각각 설치되었다.

2.2 보행 자세

SD-2 로보트의 보행을 위한 정적 보행 자세가 그림 3에 8개의 상태로 주어진다. 그림에서 사각 점선은 발이 공중에 있음을 나타내며, 발 사이의 굵은 점은 로보트의 무게중심을 나타낸다.

그림과 같이 주어진 8개의 기본 동작 자세(PP)가 로보트의 보행을 위한 각 관절의 궤적을 위한 기본 요소이다. 로보트가 보행할 때 이 8개의 PP를 차례로 지나가게 되며, 각 PP사이의 궤적은 선형 보간법에 의하여 산출된다. 로보트의 보행(좌·우 2발짝)은 1) 무게중심 이동 및 한발 지탱, 2) 발 이동 A, 3) 발 이동 B, 4) 착지 및 무게중심 복귀, 5) 무게중심 이동 및 한발 지탱, 6) 발 이동 A, 7) 발 이동 B, 8) 착지 및 무게중심 복귀의 8개의 Phase를 반복 수행하여 완성된다. 여기서 발 이동이 두 부분(2,3, 및 6,7)으로 분리된 이유는 이동되는 발이 지면과 최대로 근접할 때 발바닥과 지면의 평행 상태를 유지하여 지면과의 충돌 가능성을 배제하기 위함이다.

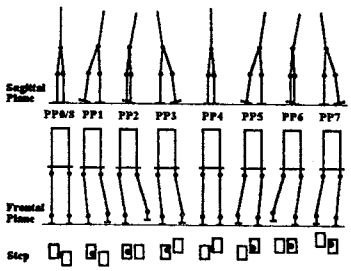

그림 3. 평지를 위한 보행 자세

보행 중인 로보트에 대한 위에서 내려다본 무게중심 변화(그림 3의 굵은 점)는 힘센서 신호로 구성된 다음 식에서 구해진다.

$$A_{cog} = L\frac{F_{lf}*1.5 + F_{lh}*0.5 + F_{rf}}{F_{tot}}, \quad (1)$$

$$B_{cog} = L\frac{F_{lf}*1.5 + F_{lh}*0.5 + F_{rf}*2 + F_{rh}}{F_{tot}}, \quad (2)$$

$$F_{tot} = F_{lf} + F_{lh} + F_{rf} + F_{rh}, \quad (3)$$

여기서 A_{cog} 및 B_{cog}는 각각 지탱하고 있는 발의 센서 값으로부터 산출되는 무게 중심을 나타내며, F의 아래 첨자 l, r, h, t, 그리고 L은 각각 좌, 우, 앞, 뒤의 센서 위치, 그리고 발의 길이를 나타낸다.

그림 3과 같이 주어진 보행 자세는 로보트가 경사면을 오를 경우 로보트의 무게 중심이 뒤로 이동되어(내리막에서는 앞으로 이동) 안정된 보행이 어렵다. 이는 보행 자세의 교정으로 극복될 수 있으며, 이를 위한 적절한 제어기를 구성하는 것이 본 연구의 목표다.

3. 신경망 제어기

이 절에서는 로보트 보행을 위한 제어기 시스템의 하드웨어 구성 및 제어 알고리즘, 즉, 신경망 및 그 학습 방법을 제시한다.

3.1 제어기의 구성

로보트 제어를 위한 전체 시스템 구성은 그림 4와 같다. 각 관절의 위치 조정을 위한 PI 제어기와 신경망으로 구성되는 궤적 발생기는 PC-486에서 C 언어로 구현되었으며, 중복기 및 센서 신호를 위하여 DDA-08 및 DAS-8의 디지털·아날로그 신호 변환기가 설치되었다. 각 관절의 모터의 구동을 위한 앰프는 단일 전력용 OP 앰프인 PA02로 설계되어 최대 *12V 2A* 까지 구동되며, 관절 위치 검출은 포텐셔미터로 하였다.

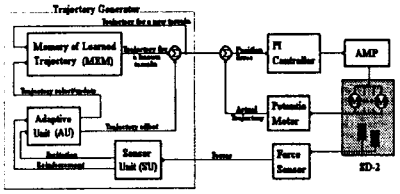

그림 4. 신경망 제어기 시스템 구성도

그림에서 센서부(SU)는 강화신호를 발생시키는 부분이며, 적응부(AU)는 신경망으로 구성되는 궤적 발생기의 핵심 부분으로 각각 다음절에서 다루어진다.

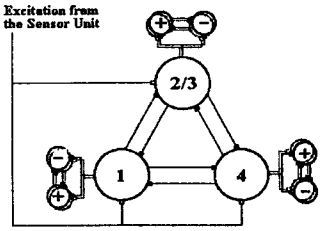

그림 5. 여기/억제 방식의 적응부 신경망

3.2 신경망 구조

그림 5는 여기/억제 방식의 신경 접속 모델[5]을 제어 대상 관절에 적용한 경우를 나타낸다. 번호가 붙은 원은 관절신경을 나타내며, 수정되어야 될 기본 보행 자세 각각에 부여된다 (총 24 신경). 즉, 측면에서 본 경사만을 고려하여 전·후 운동에 관련된 관절에 대한 신경 3개와 - 엉덩이 관절은 좌우 연동 된 동작으로 처리 - 8개의 기본 보행 자세를 위한 신경으로 구성된다. 관절신경에 붙어 있는 2 쌍으로 구성된 작은 원은 방향신경 (총 24 쌍)을 나타낸다. 그림에서 실선으로 표시된 것은 경사면을 올라갈 때(점선은 내려갈 때) 동작하는 신경을 나타내며, 원 안의 부호는 관절 각의 조절 방향을 나타낸다. 즉, "+" 신경이 동작될 때는 관절 각을 증가시키는 방향을 나타낸다. 각 원의 연결부에 있는 흰 사각형은 여기 접속(검은 사각형은 억제 접속)을 나타낸다.

이 신경망의 동작 방법을 로보트가 경사면을 올라가는 경우를 통하여 살펴보면 다음과 같다. 각 Phase가 끝나는 부분에서 센서부(SU)가 로보트의 불균형을 감지하고, 이를 적응부(AU)의 세 신경으로 여기신호를 보내면 이들 신경 중에 하나가 활성화되고 나머지 신경들은 비활성화 된다. (해당 신경 선택은 각 신경의 전 상태 값에 의해 좌우되며, 그 값들은 강화학습에 의하여 매번 수정된다.) 여기 된 관절신경(편의상 3이라 가정)은 방향신경으로 여기신호를 보내며, 앞서와 같은 방식으로 하나의 방향신경이 활성화된다. (다른 방향신경은 비활성화 됨.) 결론적으로, "+"이 활성화되었다고 생각하면,

관절 3의 각도는 로보트가 균형을 이룰 때까지 계속 증가하게 된다. 로보트가 균형을 이루면 SU에서의 여기신호가 중단되어 AU 또한 비활성화 상태가 된다. 즉 로보트는 학습된 궤적에 의하여 동작된다.

각 신경이 갖는 값을 $w(j,pp)$ $(j=1,2,3,$ 그리고 $pp=0,1,2,...,7)$라 하면, 이 값은 평면 보행 시의 힘센서 측정치와 경사면 보행 시의 측정치와의 차이로 얻어지는 강화신호에 의하여 양의 강화신호(Reward)를 최대화하고 음의 강화신호(Punishment)를 최소화하는 방향으로 매번 수정된다. 즉 현재의 자세 교정이 로보트의 균형을 향상시키면 해당 신경의 값은 증가시키고, 그 반대의 경우는 감소시킨다. 결과적으로 로보트의 균형을 향상시킨 신경은 그 값이 증가되어 계속적으로 활성화되어 로보트의 자세 교정을 더욱 가속화한다. 방향신경의 경우도 같은 학습법이 적용된다.

3.3 Reinforcement 학습 알고리즘

로보트의 균형 정도 판단은 무게중심의 변화로 가능하나 보행의 주기성을 고려할 때 다음 식으로 주어지는 힘 차이에 의함이 보다 편리하다.

$$\Delta f = f_h - f_l \tag{4}$$

여기서 f_l와 f_h는 각각 발 앞 및 뒤에서(그림 2 참조) 측정되는 힘을 나타낸다. 이 힘의 차는 로보트가 새로운 경사면에 적응한 경우 Δ_{bal}을 수평면 보행 시 힘의 차라 할 때 $\Delta_f = \Delta_{bal}$의 관계를 만족한다.

활성화될 신경은 다음 식을 만족하는 jn값에 해당하는 관절신경으로 선택된다.

$$yw(j) = w(j,pp) + n, \tag{5}$$

$$jn = \{j : \max(yw(j))\}, \quad j = 1,2,3 \tag{6}$$

여기서 pp는 현 Phase에 이용되는 PP를 나타내며, n은 구간 $[-1, 1]$에 균일하게 분포된 랜덤 변수 값으로, 각 관절신경에 대한 $w(j,pp)$가 상대적으로 비슷한 경우 각 관절에 대해 공정한 피 선택 기회를 부여하기 위해 도입된 신호이다. 방향신경에 대해서도 같은 방법이 적용되며, 여기서 선택된 관절신경 및 방향신경에 의해 궤적 변화량 $\Delta q(jn)$이 결정된다. 이 변화량에 의한 새로운 자세는 강화신호(z)로 피드백 되며,

$$\Delta f_o = \Delta f_n, \tag{7}$$

$$\Delta f_n = \Delta f_{bal} - \Delta f, \tag{8}$$

$$z = \Delta f_o - \Delta f_n, \tag{9}$$

아래와 같이 신경 강화학습에 이용된다.

$$w(jn,pp) \mathrel{+}= c * sign(s * \Delta f_n) * z, \tag{10}$$

$$s = \begin{cases} 1, & \text{if } w(jn,pp) < S \\ 0, & \text{if } w(jn,pp) \geq S \end{cases} \tag{11}$$

여기서 c는 양의 학습도 상수이며, S는 포화지수로 해당 신경 값의 유한함을 보장하기 위한 상한 값이다. 방향신경도 같은 방법으로 학습된다.

4. 실험

로보트가 경사면에 보행하는 경우 및 정지하고 있는 경우에 대한 실험을 하였다. 보행 중 학습의 경우, PP가 바뀌는 시점에서(신경 또한 바뀐다) 불연속 동작이 발생하여 로보트의 안정한 보행이 어렵다. 이 불연속 동작을 방지하기 위하여 여기서는 매 Phase의 전반부만 신경을 동작시키고 나머지 반 구간은 신경망에서 만든 관절의 변위를 점진적으로 감소시켜 다음 PP의 시작점에서 그 값을 '0'으로 만들었다.

4.1 정적 학습

로보트가 7°의 경사면 위에 PP2의 자세를 취할 때 학습되지 않은 신경에 의한 자세 교정 결과를 그림 6에 표시하였다. 관절 4의 교정 값과 신경의 값은 상대적으로 작아 지면 관계상 생략하였다.

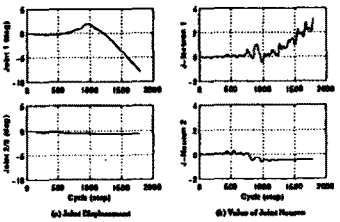

그림 6. 관절 변위와 관절신경의 값

관절신경 1의 값이 1000 스텝 이후 지배적이어서 관절 1(발목 관절)에 의한 자세 교정이 이룩되었다. 한번 학습된 신경이 계속 이용될 경우에는 그림 6과는 달리, 관절 1 만이 계속 선택되어 600 스텝 이전에 자세가 교정되는 빠른 응답 특성을 보여주었다. 특이할 사항은 로보트의 초기 자세에 따라 관절 2/3(엉덩이 관절)이 선택되어 자세 교정이 이룩되기도 하나, 두 경우 모두 경사면에 적용된 자세를 보여주었다.

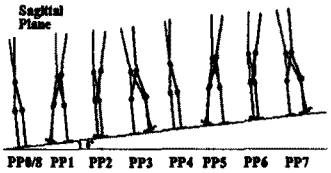

그림 7. 20회 학습 후 교정된 보행 자세

4.2 준 동적 학습

로보트가 5°의 경사면 위에서 보행하며 20회의 학습을 한 후의 교정된 보행 자세가 그림 7에 표시되었다. 그림에서 굵은 선이 교정된 자세를 나타내며, 경사면에 적응된 자세를 보여주어 제시된 제어기의 타당성을 입증하고 있다.

5. 결론

두발 로보트의 보행 자세 교정을 위한 신경망으로 구성된 제어기를 제시하였다. 제시된 제어기는 강화 학습법에 의하여 경사면에 대한 사전 정보 없이도 이를 스스로 학습하여 경사면 보행에 적합한 자세를 결정한다. 좌우 엉덩이 신경을 연 동시켜 바람직하지 않은 자세 교정을 억제시켰으며, 준 동적 보행법을 제시하여 보행 중 학습에 따른 동작의 불 연속성을 제거하였다.정적 학습 및 준 동적 학습에 대한 실험 결과를 통하여 제시된 제어기의 효용성을 보여주었다.

6. 참고문헌

[1] Y.F. Zheng and F. Sias, "Design and motion control of practical biped robots," Int. J. of Robotics and Automation, vol.3, no.2, pp. 70-78, 1988.

[2] Y.F. Zheng and J. Shen, "Gait synthesis for the SD-2 biped robot to climb sloping surface," IEEE Trans. on Robotics and Automation, vol. 6, no. 1, pp. 86-96, 1990.

[3] E.R. Dunn and R.D. Howe, "Towards smooth bipedal walking," Proc. 1994 IEEE Int. Conf. on Robotics and Automation, San Diego, CA, 1994, pp. 2489-2494

[4] V. Gullapalli, J.A. Franklin, and H. Benbrahim, "Acquiring robot skills via reinforcement learning," IEEE Control Systems, pp. 13-24, Feb. 1994.

[5] A.G. Barto, R.S. Sutton, and C.W. Anderson, "Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems," IEEE Trans. on Sys., Man, and Cyber., vol. SMC-13, no. 5, pp. 834-846, 1983