

Topological Map을 이용한 이동로봇의 행위기반 학습제어기

이석주*†, 문정현†, 한신‡, 조영조†, 김광배†
 † 한국과학기술연구원 지능제어연구센터, ‡ 연세대학교 전기공학과, ¶ 고려대학교 전기공학과

Behavior-based Learning Controller for Mobile Robot using Topological Map

Seok-Joo Yi*†, Jung-Hyun Moon†, Shin Han‡, Young-Jo Cho†, Kwang-Bae Kim†
 † ISCRC, KIST, ‡ Dept. of Electrical Eng., Yonsei Univ., ¶ Dept. of Electrical Eng., Korea Univ.

Abstract - This paper introduces the behavior-based learning controller for mobile robot using topological map. When the mobile robot navigates to the goal position, it utilizes given information of topological map and its location. Under navigating in unknown environment, the robot classifies its situation using ultrasonic sensor data, and calculates each motor schema multiplied by respective gain for all behaviors, and then takes an action according to the vector sum of all the motor schemas. After an action, the information of the robot's location in given topological map is incorporated to the learning module to adapt the weights of the neural network for gain learning.

As a result of simulation, the robot navigates to the goal position successfully after iterative gain learning with topological information.

1. 서 론

자율 이동로봇을 미지의 환경에서 주행할 수 있도록 제어하는 것은 매우 힘든 작업이다. 예를 들면, 정확하지 않은 센서 데이터, 센서의 실시간 계측, 센서 정보에 의한 외부 환경의 모델링 등 수 많은 문제가 발생할 수 있다. 특히 초음파와 센서의 경우 초음파의 특성에 의해 약간의 각도차에도 측정값은 크게 변하고, 다른 센서에 의해 발생된 초음파끼리 간섭을 일으키는 등 정확한 거리정보를 얻기 힘들다.

이러한 문제를 해결하기 위한 방안으로서 이동로봇에 "지능"을 부여하고자 하는 연구가 이동로봇 뿐만 아니라 제조 시스템, 제어, 화상처리, 그리고 통신 등 다양한 분야에서 큰 줄기를 이루어오고 있다.

지능이라는 개념은 정의하기가 매우 난해하고 연구자들에 따라 서로 다른 용어를 포함하기도 하지만 [1], [2] 보편적으로 "학습"을 지능 시스템의 가장 큰 특징으로 분류하고 있다.

또한 이동로봇을 대상으로 할 때 지능 시스템에 대한 또 한가지의 중요한 특징으로서 외부 환경의 변화에 대한 생물의 즉각적인 행위(behavior)를 들 수 있다. 즉, 이동로봇의 모든 작업은 하나의 큰 목적 아래 매 시간에 대한 환경 변화에 따라 수행되는 일련의 행위들로 구성된다. 이러한 행위기반 제어기법에는 여러 가지 방법이 있지만 그 중에서도 센서정보에 의해 즉각적인 행위를 수행하는 모터 스키마(schema) 방법이 가장 널리 연구되고 있다[3].

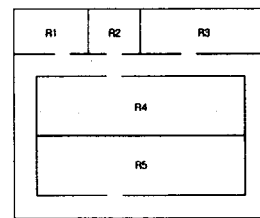
본 논문에서는 이동로봇에 지능을 부여하는 한 방법으로서 행위기반 학습 제어기를 제안하였다. 이동로봇이 실내에서 특정 목적지까지 주행하는 것을 하나의 임무로 설정하였다. 작업 영역에 대한 외부 환경은 위상 지도(topological map)를 이용해 표현하고, 이러한 지도

정보를 이용해 주어진 목적지까지 주행하는 동안 마주치는 다양한 주위 환경에 대해 수행된 자신의 행위에 대한 결과를 바탕으로 다음 행동을 위한 각 모터 스키마의 이득값을 조정하게 된다. 제안된 제어기의 성능을 모의실험을 통해 검증하였다.

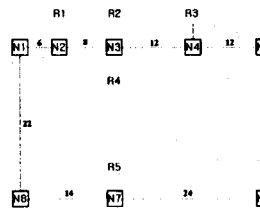
2. 외부 환경의 표현

이동로봇이 주어진 임무를 수행하기 위해서는 반드시 외부 환경과의 상호 작용이 필요하다. 이를 위해서는 외부 환경으로부터 현재 자신의 위치 정보와 목적지에 대한 위치 정보를 인식할 수 있어야 한다. 일반적으로 외부 환경의 표현은 수집된 센서정보를 통해 지도를 작성한 형태로 표현할 수 있다. 이동로봇의 지능이라는 측면을 생각한다면 스스로 주행 중 수집된 센서정보를 이용하여 지도를 작성해야 할 것이다. 본 논문에서는 사용자가 작성한 지도 정보를 이동로봇에 부여하고 이 지도정보를 통해 현재 자신의 위치를 파악할 수 있다고 가정하였다.

외부 환경 지도는 초음파와 센서 등에 의해 수집된 센서정보를 이용하여 이동로봇의 활동 영역을 격자 셀로 분할해서 각 셀이 점유되어 있는지를 표현하는 격자 지도 [4] 방법과 외부 환경의 특징점들을 추출하여 거리정보와 방향, 그리고 연결정보를 표현하는 위상 지도[5] 등 여러 방법이 있지만 본 논문에서는 위상 지도를 이용하였다. 그림 1은 모의 실험을 위한 이동로봇의 외부 주행 환경과 그에 대한 위상 지도를 나타낸다.



(a) 외부 주행환경



(b) 위상 지도

그림 1. 이동로봇의 외부 주행환경의 표현

3. 행위기반 학습 제어기

3.1 행위기반 학습제어기의 구조

이동로봇은 목적지점까지 주행하는 동안 주위 외부환경과 끊임없이 상호작용을 통해 정보를 수집하고 이를 통해 다음 행위를 결정해야 한다. 따라서 이동로봇의 제어기는 외부환경으로부터 정보를 수집하는 인식모듈과 각 행위에 대한 모터 스키마를 결정하는 행위모듈을 반드시 포함해야 한다.

제한된 행위기반 학습제어기는 이러한 인식모듈, 행위모듈과 결정된 각 스키마들을 조정하여 최종적으로 이동로봇의 구동 모터에 전달하는 행위조정모듈과 한 번의 동작 후 그 결과를 토대로 각 스키마에 대한 이득을 학습하는 학습모듈로 구성된다. 마지막으로 이런 모든 모듈들을 관리하여 전체적인 임무를 수행하도록 감독하는 관리모듈이 가장 상위에 구성된다.

인식모듈을 통해 외부환경에 대한 정보가 수집되면 행위모듈에서는 이 정보를 바탕으로 각 행위에 대한 모터 스키마를 계산하고 학습모듈에서 학습된 이득값을 적용하여 모터의 구동 벡터를 구하게 된다. 이렇게 하나의 동작이 이루어진 후 다시 수집된 센서정보를 이용하여 현재 위치에서의 위상지도 정보와 함께 학습모듈에서 새로운 이득값을 학습하게 된다.

따라서 위와 같은 과정을 반복하여 이동로봇의 각 행위에 대한 학습이 이루어지게 된다.

그림 2는 제한된 행위기반 학습제어기의 구조와 외부환경과의 상호작용을 보여준다.

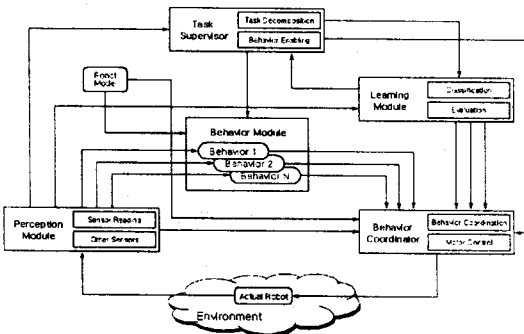


그림 2. 행위기반 학습제어기 구조

3.2 각 행위의 조정

모터 스키마 방법은 모든 행위들에 대한 각각의 모터 스키마를 결정하고 각 스키마 벡터들을 합산하여 모터의 구동 벡터를 결정한다. 이 방법의 단점은 현재 상황 또는 현재의 외부환경에서는 필요하지 않은 행위들에 대해서도 스키마들을 계산하고 결과에 반영시킨다는 것이다. 상황에 따라서는 특정 행위가 전혀 필요하지 않을 수도 있고, 또는 여러 가지 행위들이 필요하지만 그 중에서도 특정 행위가 가장 중요하게 고려되어야 할 경우도 있다. 따라서 상황에 맞게 각 행위들의 중요성을 고려하기 위해 각 행위에 대한 스키마에 이득값을 적용한 이후에 벡터 합산을 수행하도록 하였다.

목적지점까지의 주행이라는 작업을 위해 사용된 스키마들은 다음과 같고, 그림 3은 이러한 행위 조정방법을 보여준다.

- Move_to_Goal : 로봇으로부터 목적지점까지의 인력(attractive force)
- Follow_Wall : 양쪽 벽면 중앙으로의 주행
- Avoid_Obstacle : 장애물 회피
- Detour : 과거 행위를 기억하여 Local minima 탈출

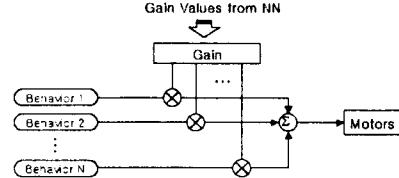


그림 3. 각 행위의 조정

3.3 이득값의 학습

3.3.1 행위 학습의 개념

이동로봇에 있어서 행위의 학습이라는 개념은 먼저 행위라는 개념에서부터 출발한다. 행위는 이동로봇의 가장 기본적인 동작을 의미하며, 센서로부터 들어오는 정보에 대해 즉각적으로 응답하는 reactive response의 개념으로 이해해야 하며 이러한 행위의 학습은 이렇게 수행된 각 행위의 결과가 외부환경에 미치는 영향까지 고려하여 이루어져야 한다. 즉 행위 그 자체는 학습할 수 있는 대상이 아니며 각 상황에서의 행위와 그 결과를 학습 대상으로 해야 한다. 그림 4는 이와 같은 행위와 행위 학습에 대한 개념을 보여준다.

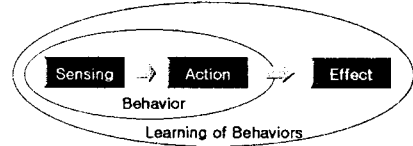


그림 4. 행위 학습의 개념

3.3.2 학습 방법

행위의 학습은 현재 상황에 대한 각 행위의 이득값을 신경회로망을 통해 학습하게 된다. 학습모듈은 외부로부터 강화 신호(external reinforcement signal)(6)를 받아들이며 학습을 수행한다. 강화학습의 장점은 학습에 임출력 교사 데이터가 필요한 기존의 역전파알고리즘(feedforward error back-propagation algorithm)과 달리 외부의 강화신호에 의해 신경회로망의 학습이 이루어지므로 교사 데이터가 필요하지 않다는 점이다. 이에 이용된 신경회로망은 입력층, 평가층, 출력층의 3개 층으로 구성되지만 일반적인 역전파알고리즘을 이용한 신경회로망의 입력층, 은닉층, 출력층의 구조와는 다른 개념이다.

학습과정을 살펴보면 먼저 수집된 센서정보로부터 현재의 상황을 분류하여 신경회로망의 입력뉴런을 결정한다. 따라서 결정된 입력뉴런만이 1의 값을 갖도록 했으며 나머지 입력뉴런에는 0이 전달되도록 하였다. 이 입력뉴런에 연결된 연결 하중값과 이전 행위의 결과에 의해 위상 지도와의 연관관계에 따라 외부 강화신호가 주어지면 각 행위에 대한 이득값이 출력되는 구조로 학습이 이루어진다. (그림 5)

각 행위에 대한 현 상황에서의 이득값은 다음과 같이 결정된다.

$$g_i = f \left(\sum_{k=0}^m I_k \cdot W_{ki} + \sigma(t) \right) \quad (1)$$

이때 g_i 는 i 번째 행위의 이득, I_k 는 k 번째 입력값, W_{ki} 는 입력층과 출력층 사이의 연결 하중값이다. 그리고 $\sigma(t)$ 는 $[-1, 1]$ 사이의 임의의 값, m 은 입력뉴런의 개수이다. 그리고 함수 f 는 다음과 같이 정의되는 sigmoid 함수이다.

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (2)$$

연결 하중값 W_{ki} 는 다음과 같이 갱신된다.

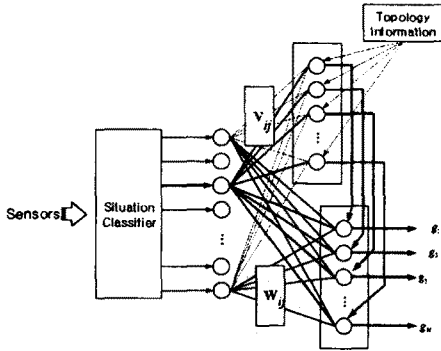


그림 5. 학습 모듈의 신경회로망 구조

$$W_k(t+1) = W_k(t) + \alpha \cdot b_k(t) \cdot e_k(t) \quad (3)$$

이때, α 는 학습률, $b_k(t)$ 는 내부 강화신호, $e_k(t)$ 는 식 (4)와 같이 주어지는 적임도(eligibility)를 의미한다.

$$e_k(t+1) = \delta \cdot e_k(t) + (1-\delta) \cdot I_k \cdot a_k \quad (4)$$

δ 는 [0,1] 사이의 값이다. 내부 강화신호 $b_k(t)$ 는 신경 회로망의 또 다른 층인 평가층에 의해 주어진다. 평가층에서는 내부 강화신호의 예측값 $P_i(t)$ 을 식 (5)와 같이 계산한다.

$$P_i(t) = \sum_{k=0}^n V_k \cdot I_k \quad (5)$$

V_k 는 입력층과 평가층 사이의 연결 하중값이다. 내부 강화신호 $b_k(t)$ 는

$$b_k(t) = r(t) + \gamma \cdot P_i(t) - P_i(t-1) \quad (6)$$

과 같이 계산된다. 이때 γ 는 [0,1] 사이의 값이다. 그리고 V_k 는 다음과 같이 갱신된다.

$$V_k(t+1) = V_k(t) + \beta \cdot b_k(t) \cdot X_k(t) \quad (7)$$

이때 β 는 [0,1]의 상수값이고, X_k 는 평가층의 적임도이다.

$$X_k(t+1) = \lambda \cdot X_k(t) + (1-\lambda) \cdot I_k(t) \quad (8)$$

4. 모의실험 및 결론

외부 강화신호 $r(t)$ 는 하나의 동작이 끝난 후 위상 지도에 의한 정보로부터 등급화된 값으로 지정했다. 즉, 벽과의 충돌이 없으면 0, 충돌이 발생하면 -1로 지정된다. 그리고 이동로봇이 만일 복도에 위치하고 있다면 여러 가지의 행위들 중에서 Follow Wall 행위에 대한 외부 강화신호를 +1로 할당하는 등 위상 지도상의 특징에 의해 외부 강화신호를 지정하였다.

목적지점까지 주행하는 임무에 대해 모의실험에 사용된 각 상수값들은 다음과 같다.

표 1. 학습모듈에 사용된 상수값

상수	값
α	3
δ	0.3
γ	0.6
β	0.7
λ	0.4

또한 외부 강화신호 $r(t)$ 는 다음과 같이 할당하였다.

표 2. 위상 지도 정보를 이용한 외부 강화신호

각 동작 후 상황	스키마	$r(t)$
Corridor	Move to Goal	0.5
	Follow Wall	1
	Avoid Obstacle	1
Hall	Detour	0.2
	Move to Goal	1
	Follow Wall	0.5
	Avoid Obstacle	1
Collision	Detour	0.5
		-1

신경회로망의 모든 연결 하중값들은 초기에 모두 임의로 선정되었고, 이러한 초기값으로부터 반복적으로 주행을 하여 학습시킨 결과 다음과 같이 N1 지점에서 N6지점까지 안정적으로 주행하는 것을 볼 수 있었다.

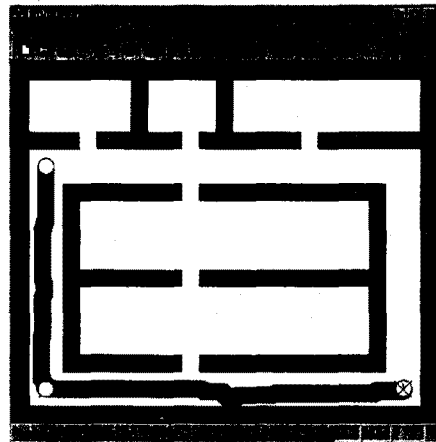


그림 6. 학습을 통한 목적지점까지의 주행(모의실험)

앞으로 위상 지도를 로봇 스스로 작성하는 방법과 자기 위치 판별에 대한 연구가 수행되어야 할 것이다.

(참 고 문 헌)

- [1] W. Fritz et. al., "The Autonomous Intelligent system", *Robotics and Autonomous Systems*, Vol. 5, pp. 109-125, 1989
- [2] A. Falli, "Integrating Learning and Decision-Making in Intelligent Manufacturing Systems", *Intelligent and Robotics Systems*, Vol. 3, pp. 117-130, 1990
- [3] R. C. Arkin, "Motor Schema-Based Mobile Robot Navigation", *The International Journal of Robotics Research*, Vol. 8, No. 4, pp. 92-112, 1989
- [4] J. Borenstein, et. al., "The Vector Field Histogram - Fast Obstacle Avoidance for Mobile Robots", *IEEE Trans. on Robotics and Automation*, Vol. 7, No. 3, pp. 278-288, 1991
- [5] B-S Ryu and H-S Yang, "Integration of Reactive Behaviors and Enhanced Topological Map for Robust Mobile Robot Navigation", *IEEE Trans. on Systems, Man, and Cybn.*, Vol. 29, No. 5, pp. 474-485, 1999
- [6] R. S. Sutton and A. G. Barto, "Reinforcement Learning", The MIT Press, 1998