지능로봇을 위한 행동선택 및 학습구조

An Action Selection Mechanism and Learning Algorithm for Intelligent Robot

윤 영 민 · 이 상 훈 · 서 일 홍
Young Min Yoon · Sanghoon Lee · Il Hong Suh

Abstract - An action-selection-mechanism is proposed to deal with sequential behaviors, where associations between some of stimulus and behaviors will be learned by a shortest-path-finding-based reinforcement learning technique. To be specific, we define behavioral motivation as a primitive node for action selection, and then sequentially construct a network with behavioral motivations. The vertical path of the network represents a behavioral sequence. Here, such a tree for our proposed ASM can be newly generated and/or updated, whenever a new sequential behaviors is learned. To show the validity of our proposed ASM, some experimental results on a "pushing-box-into-a-goal task" of a mobile robot will be illustrated.

Key Words : action-selection-mechanism, reinforcement learning, behavioral motivation

1. 서론

지능 로봇을 구현하기 위해서는 외부 환경과 내부 환경을 조합하여 가장 적절한 행동을 선택할 수 있는 행동 선택 구조와 변화하는 환경에 적응할 수 있는 학습 방법이 필요하다. 행동 선택 구조를 실제 이동 로봇에 적용하는 경우 실제 센서로부터 들어오는 데이터를 통한 외부 상태의 인식이 문제로 나타날 수 있다. 본 논문에서는 지능 로봇을 위한 행동 선택 구조 및 학습 방법을 실제 이동 로봇에 적용해 봄으로써 적용 시 나타나는 문제점과 그에 대한 해결책을 제시한다. 행동 선택 구조는 계층 구조로 이루어져 있으며 주어진 상황에서 적절한 임무를 선택할 수 있고 그 임무를 만족시키기 위해 적절한 행동을 선택할 수 있다. 제안된 구조의 타당성을 검증하기 위해 실제 이동 로봇에 적용하여 실험하고 실제 센서로부터 들어오는 외부상태의 인식을 보완하기 위해 Bayesian Filter를 적용하였다.

2. 행동선택구조

2.1 행동동기

제안된 행동선택구조는 행동동기(BM : Behavioral Motivation)로 이루어진 계층구조이다.

행동 동기는 행동선택의 기본 단위로서 기본구조는 그림 1과 같으며 지능로봇의 내부 상태와 외부 자극의 영향을 조합하여 행동을 선택하는데 기준이 되는 값을 계산한다.

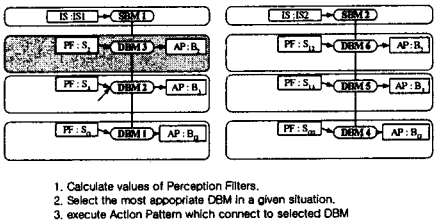

그림 1. 행동동기 기본구조

제안된 구조에서 행동동기들은 두 가지 타입으로 분류된다. 첫 번째 타입은 정적 행동동기(SBM : Static Behavioral Motivation)로 동기 또는 임무를 나타낸다. 두 번째 타입은 동적 행동동기(DBM : Dynamic Behavioral Motivation)로서 정적 행동동기를 만족시키기 위한 순차적인 행동들로 구성 및 학습될 수 있다.

SBM과 DBM에 영향을 주는 요소는 자극인식필터(PF : Perception Filter), Releaser가 있다. PF는 입력 센서로부터 들어오는 자극, 이벤트들 중에서 관련된 자극이나 이벤트만을 식별하고, 관련 정도를 출력하는 역할을 한다. Releaser는 행동동기를 계층구조로 구성할 경우 하위 단계의 행동동기의 값이 결정되기 위해서는 상위 노드로부터 센서정보가 막히지 않고 하위 노드까지 전달되어야 한다. 이를 위해 releaser라는 자극인식필터들을 이용하여 sensory bottleneck 현상을 방지한다.

2.1.1 정적 행동동기(SBM)

정적 행동동기는 임무 또는 동기를 나타낸다. SBM들의 값은 주어진 내 · 외부 상태에서 어떤 동기를 활성화 시킬

저자 소개
* 윤 영 민 : 漢陽大學 情報通信學科 碩士課程
** 이 상 훈 : 漢陽大學 電子電氣制御計測學科 博士課程
***서 일 홍 : 漢陽大學 情報通信學科 助教授 · 工博

지를 결정하는데 사용되며 계산하는 식은 식(1)과 같다.

$$V_{SBM_{i(t+1)}} = \sum V_{IV_{ij}} + \sum V_{PF_{ik}} - \sum_{allSBM} (V_{SBM_{il}} I_{il}) + effect^{i}_{DBM} \quad (1)$$

i : index of where the i th SBM
j : index of related IS
k : index of related PF
I_{il} : inhibitory Gain that SBM_l applies a inhibition
l : index of same level SBM
$effect^{i}_{DBM} = w_i V_{DBM_m}$
m : index of maximum valued DBM
w_i : weight of feedback value from DBM under the i th SBM

각각의 SBM 들은 동적 행동동기 그룹을 가지고 있으며 이 DBM 그룹은 해당 SBM을 만족할 수 있는 연속적인 DBM들로 구성되어 있다. 그림 2는 이러한 SBM들이 사전에 구성된 예를 보여준다.

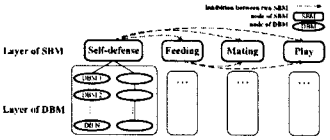

그림 2. SBM, DBM 구성 예

2.1.2 동적 행동동기(DBM)

골(goal) 또는 임무를 수행하기 위해 지능 로봇은 일련의 행동들을 만들어야 하고 그 중에 가장 적절한 행동을 선택해야 한다. 이런 이유로 동적 행동동기들은 학습한 내용에 따라 유연하게 변할 수 있는 계층구조로 구성된다. 각 DBM의 값은 자극인식 필터(PF : Perception Filter), 부모 BM, releaser들의 값에 의해 계산되고 자식 BM과 연관된 자극이 들어 왔을 때, 계산된 값은 다시 자식 BM으로 보내진다. DBM의 구조는 그림 3과 같다.

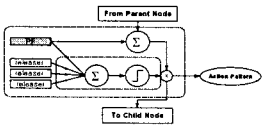

그림 3. DBM 구조

DBM의 값은 수직 경로를 따라 누적되며 releaser는 그 값의 흐름을 봉쇄하는 역할을 한다. DBM을 식(2)와 같이 계산하고 주어진 상황에서 가장 적절한 DBM은 그룹에서 가장 높은 값을 선택하는 것으로 식(3)과 같다.

$$V_{DBM_i} = (V_{DBM_{i-1}} + V_{PF_i}) STEP\left(\sum_{k=1}^{m} V_{Releaser_k}\right) \quad (2)$$

i : index of this node
j : index of related PF
k : index of related releaser

$$STEP(x) = \begin{cases} 1, & \text{for } x > 0 \\ 0, & \text{for } x = 0 \end{cases}$$

$$selectedDBM = \underset{i \in allDBMunderSBM}{\arg\max} (DBM_i) \quad (3)$$

2.2 행동선택 절차

제안된 행동선택 구조는 매번 주어진 내·외부 상태에 따라 가장 적절한 행동을 선택해야 하며 이를 위해 관련된 SBM, DBM을 선택한다. SBM과 DBM을 선택하는 절차는 다음과 같이 요약할 수 있다.

1. 각 DBM 그룹에서 가장 높은 값을 갖는 DBM 선택
2. 식 (1)에 의해 각 SBM이 계산되고 가장 높은 값을 갖는 SBM이 선택

그림 4는 제안된 행동선택 구조의 전체 과정을 보여준다.

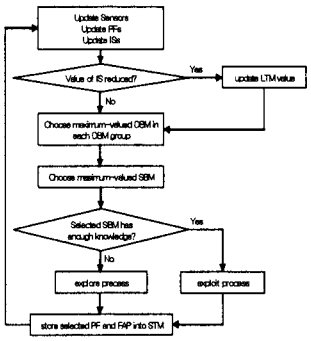

그림 4. 행동선택 절차

3. 학습구조

3.1 학습절차

학습을 위해 최단 경로 찾기 알고리즘이 포함된 강화학습 알고리즘이 적용되었다. 우선 강화학습은 성공적으로 임무를 수행했을 때 적용된다. 이때, 단기기억(STM : Short Term Memory)에 저장됐던 모든 S-R(Stimulus-Response) 연관들이 장기기억(LTM : Long Term Memory)으로 전환된다. 그 후 초기 S-R연관에서 골까지의 최단경로를 찾기 위해 최단경로 찾기 알고리즘이 적용된다. 최단 경로 내에 있는 S-R 연관들은 낮은 신뢰도를 할당 받는다. S-R 연관들이 LTM에 기록된 신뢰도를 기반으로 선택되기 때문에, 이렇게 함으로써 잘못된 행동을 선택할 확률을 줄일 수 있고 학습속도 또한 높일 수 있다.

내부 메모리는 STM과 LTM으로 구성되며 그림 5는 STM에 저장된 S-R 연관들이 LTM으로 전환되는 과정을 보여준다.

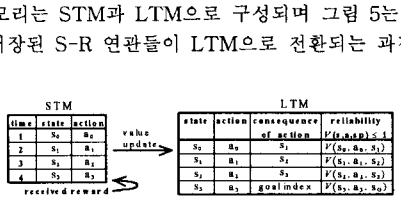

그림 5. STM에서 LTM으로의 전환

일단 STM에 있는 모든 S-R 연관들이 LTM으로 전환되면 다음 탐색을 위해 STM에 있는 모든 S-R 연관들은 지워진다. STM에 있는 모든 S-R 연관은 보상을 받기 위한 사전

행동으로 볼 수 있다. 보상을 받은 시점에서 시간적으로 가까운 S-R 연관은 시간적으로 먼 S-R 연관에 비해 중요도가 크기 때문에 높은 신뢰도를 가져야 한다. 신뢰도의 값은 식(4)에 의해 결정되며 이는 강화학습 방법 중 몬테-카를로 방법(Monte-Carlo method)과 유사하다.

$$V(s_t, a_t, s_{t+1}) \leftarrow V(s_t, a_t, s_{t+1}) + \frac{\eta(1 - V(s_t, a_t, s_{t+1}))}{i^{\lambda}} \qquad (6)$$

s_t : the index of the i^{th} stimulus
a_t : the index of response behavior
η : learning rate
λ : decay rate
i^{λ} : weightings of distance from reference time

4. 구현 및 실험

4.1 실험환경

제안된 행동선택 구조를 검증하기 위하여 2m×2m의 공간에서 로봇이 상자를 밀어 골까지 가는 실험을 하였다. 로봇은 이동과 회전의 동작을 할 수 있으며 외부로부터 정보를 받는 센서는 vision 만을 사용하였다. 센서로부터 들어오는 정보인식의 오차극복을 위해 Bayesian Filter를 적용하였다. 학습해야 할 규칙은 다음과 같다.

1. move : 로봇과 상자가 일직선이 되면 보상
2. turn : 로봇과 상자와 골 모두가 전방에 있으면 보상
3. push : 상자를 밀어 골에 가까워지면 보상

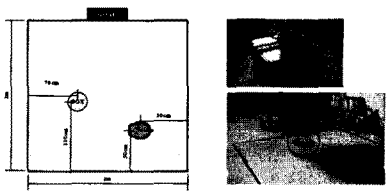

그림 6. 실험환경

4.2 Bayesian Filter의 적용

행동 선택 구조를 실제 이동 로봇에 적용하는 경우 실제 센서로부터 들어오는 데이터를 통한 외부 상태의 인식이 문제로 나타날 수 있다. 이러한 문제점을 보완하기 위하여 본 실험에서는 Bayesian Filter 방식 중 grid-based approach를 이용하였다. 이 방식은 10cm에서 1m 크기 사이의 small patches 안에 격자구조의 바닥을 만들어 위치를 추정하는 방식으로 각각의 grid cell안에 robot 혹은 obhect에 대한 belief를 포함한다. 업데이트 방식은 기본적인 bayes filter update equations에서 적분 대신에 덧셈의 방법을 사용하여 업데이트 한다. 이 방식의 장점은 일시적으로 불연속적인 분포구역이 나타나더라도 해결이 가능하며 sensor noise가 높은 환경에서도 비교적 정확한 위치판단이 가능하다. 따라서 본 실험에 적합한 방식이라 판단, 적용하였다.

4.3 실험결과

실험이 수행된 후 계층구조에는 보상을 받기 위한 행동동기들이 추가되었음을 볼 수 있었다. 또한 제안된 행동선택 구조가 실제 로봇에도 적용될 수 있음을 확인할 수 있었다. 아래 표는 실제 로봇에 제안된 구조를 적용하여 실험한 결과 학습이 성공하였음을 보여주고 있다.

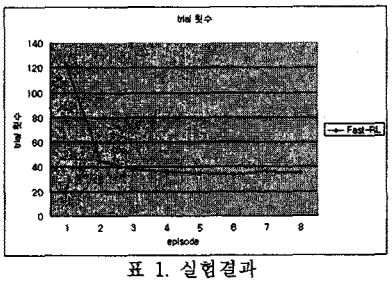

표 1. 실험결과

5. 결론

주어진 환경에서 살아남기 위해 적절한 행동 선택과 적절한 행동을 학습하는 것은 지능 로봇에 있어 가장 중요한 능력이다. 이를 위해 본 논문에서는 정적 행동동기와 동적 행동동기로 구성된 행동선택 구조 및 학습방법과 이를 실제 로봇에 적용 시 나타나는 외부 상태 인식문제에 대한 해결방법을 제시하였다.

실험 시에 실제 센서로부터 들어오는 데이터의 낮은 신뢰도 때문에 현재 로봇이 처한 상태를 파악하기 힘들고, 이로 인해 학습 및 적절한 행동 선택을 하기가 어려운 점을 극복하기 위해 Bayesian Filter를 적용하였다. 이를 통해 제안된 행동선택 및 학습구조가 실제 로봇에 적용될 수 있음을 확인하였다.

참 고 문 헌

[1] Bruce M Blumberg, *"Old Tricks, New Dogs"*, Ethology and Interactive Createures,1997

[2] P. Maes.,*"Modeling Adaptive Autonomous Agents."*,Artificial Life Journal, edited by c. Langton, Vol. 1, No. 1 & 2, pp.135-162, MIT press, 1994.

[3] R. Brooks, *"A Robust Layered Control System For a Mobile Robot."*, In IEEE journal of Robotics and Automation, pages 14 23, April, 1986.

[4] Lorenz, K. *Foundations of Ethology.*, Springer-Verlag, New York, 1973

[5] Toby Tyrrell *"Computational Mechanism for Action Selection"* , Ph.D. Thesis, Centre for Congitive Science, University of Edinburg, 1993

[6] R. Sutton, A. Barto, *Reinforcement Learning*, MIT Press, 1997.

[7] N.Tinbergen, *The Study of Instinct.*,Oxford University Press., 1951