# 지능형 소프트웨어 로봇을 위한 행동학습구조

°권우영*,민현석*,장국현*,이상훈***,서일홍*
* 한양대학교 정보통신대학원 정보통신공학과 (Tel : +82-2-2290-0392;E-mail:ihsuh@hanyang.ac.kr)
** 한양대학교 전자전기제어계측공학과(Tel : +82-31-408-5802;E-mail:ihsuh@hanyang.ac.kr)

# Behavior Learning Architecture for Intelligent Software Robot.

°Woo-Young Kwon*, Hyun-Suk Min*, Zhang Guo-Xuan*, Sang-Hoon Lee*, Il Hong. Suh*
* The graduate school of information & communications , Hanyang University
(Tel : +82-2-2290-0392;E-mail:ihsuh@hanyang.ac.kr)
** School of Electrical Engineering and Computer Science, Hanyang University
(Tel : 82-31-408-5802;FAX:82-31-408-5803;E-mail:ihsuh@hanyang.ac.kr)

**Abstract** - 기존의 로봇은 주로 예측 가능한 환경 하에서 동작해왔다. 그러나 로봇의 적용분야가 확대되면서 예측하기 힘든 복잡한 자극에 대해 반응하도록 요구되고 있다. 복잡한 자극은 동일시간에 여러 가지 자극이 존재하는 공간적 복잡성과, 각기 다른 시간에 자극이 연속적으로 배열된 시간적 복잡성을 가진다. 기존의 로봇은 복잡한 자극에 대한 대처능력이 취약하다.

이러한 환경에서 적응할 수 있도록 여러 방면의 연구가 진행되어 왔으며, 그 중에서 동물이 환경의 변화에 대처하는 방법에 관한 많은 연구들이 진행되고 있다.

본 논문에서는 시간적 복잡성을 가진 자극에 반응하고 이를 학습하기 위해 HMM(Hidden Markov Model)을 이용한 시계열 학습구조를 제안한다. 또한 기본적인 행동선택 및 학습을 위해 동물의 행동선택을 모델링한 구조를 구현하였다.

## 1. 서 론

로봇은 주어진 환경에서 환경에 관한 정보를 습득하고 목적을 성취하기 위하여 최적의 행위를 선택하는 능력을 소유하여야 한다[1]. 그러나 지난 10년 동안의 연구에도 불구하고 로봇에 대한 연구는 제한적인 지능과 단순한 작업을 수행하는 단계였다. 또한 동적으로 변하는 환경에서 더욱 제한적인 행동 영역과 지능의 한계를 가지고 있다.

이러한 문제를 해결하기 위해 로봇은 행동의 선택(Behavior Selection)에 있어서 내부 상태(Internal State)와 외부자극에 대해 적절한 행동을 취해야 하고, 과거의 경험을 바탕으로 하여 동적으로 변화하는 환경에 적응할 수 있는 지능을 가지고 있어야 한다. 또한 단순한 자극뿐 아니라 연속적인 패턴에 대하여 예측하거나 행동할 수 있는 절차적 기억(Procedural Memory)을 가지고 있어야 한다[2][3].

본 논문에서는 지능형 소프트웨어 로봇을 설계하고 구현하여 앞에서 설명한 문제를 해결하고자 하였다. 즉 주어진 목적과 환경에 있어서 정보를 습득하고 패턴을 인식하기 위한 센서 모듈, 이러한 외부자극을 적절하게 행동과 연결시킬 수 있는 Release Mechanism 모듈, 자극과 행동을 기억하고 이를 학습하기 위한 기억 모듈, 외부자극과 내부상태 그리고 과거의 기억을 바탕으로 가장 적절한 행동을 선택할 수 있도록 하는 행동모듈, 내부 상태에 따른 의도적인 행동을 하기 위한 방법으로 내부상태 모듈을 설계, 구현하였다.

특히 절차적 기억을 구현하기 위하여 HMM을 이용하였다. 소프트웨어 로봇이 동적으로 들어오는 연속적인 패턴에 대하여 인지하고 그 다음 패턴을 예측하여 가장 적절한 행동을 할 수 있는 HMM으로 고차원적인 학습 및 연속적인 패턴에 대한 인식이 가능한 에이전트를 설계, 구현하였다.

## 2. 행동 학습 구조와 행동 선택 과정

전체 시스템의 구성은 크게 행동선택 엔진부와 가상환경으로 구현되어 있으며, 전체 구조는 그림 1과 같다.

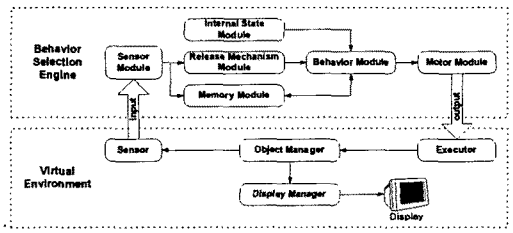


그림 1 전체구조

### 2.1 행동선택 엔진(Behavior Selection Engine)

a) 센서 모듈(Sensor Module)

센서모듈의 기능은 외부 환경에서 들어온 자극을 엔진에서 처리할 수 있도록 변환하여 RM 모듈로 전달한다(그림 2). RM 모듈로 전달되는 정보는 object의 ID와 object와의 거리이다. 이는 엔진의 입력단에 해당한다.

b) Release Mechanism 모듈

RM 모듈(Release Mechanism Module)은 엔진 내부에 저장된 외부자극에 대한 정보를 필터링하고, 가중치를 학습하여 행동단위에 그 값을 전달하는 역할을 담당한다[4].

RM은 입력된 자극과 행동간의 관계를 결정짓는다. RM은 어떠한 자극을 통과시킬지(object matching stage), 통과된 자극의 값이 어떻게 결정되는지(weight stage), 통과된 자극이 어떤 행동에 영향을 미칠지(behavior connection)의 세 가지 요소로 구성된다. RM의 구조는 그림 2와 같다.

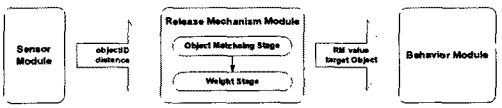


그림 2 행동선택 엔진에서의 신호흐름

c) 내부상태 모듈(Internal State Module)

내부상태 모듈은 감정 및 행동을 유발시키는 동기를 담당하게 된다.

내부상태 값은 다음 식에 의해 갱신된다[5] .

$$IS_{it} = (IS_{i(t-1)} \cdot damping_i) + growth_i - \sum_k effect_{kit}$$

$IS_{it}$ =시간 $t$에서 내부상태 $i$의 값
$damping_i$ =내부상태 $i$의 감쇄비율
$growth_i$ =내부상태 $i$의 증가비율
$effect_{kit}$ =$modifyGain_{ki} \cdot Value_{k(t-1)}$
$modifyGain_{ki}$ =행동단위 $k$가 내부상태 $i$에 영향을 주는 정도
$Value_{k(t-1)}$ =시간 $t-1$에서 행동단위 $k$의 값

내부상태 모듈은 보상을 받을 수 있는 특정한 행동모듈(예 : 먹는 행동)과 연관되어 있다. 연관된 행동이 수행되면 그 값에 비례해서 내부상태 값이 감소하게 된다. 내부

상태 값이 감소하는 현상은, 연관된 행동에 대해 만족을 느낀다는 의미이며, 이는 보상을 받은 것으로 간주된다. 따라서 내부상태의 감소되는 시점은 학습이 이루어지는 시점이 된다. 내부 상태 값은 매 시간 갱신되며 이는 전 상태의 값과 damping, growth, 그리고 외부효과(effect)의해 영향을 받게 된다.

d) 행동모듈(Behavior Module)

행동모듈은 각각의 행동단위로 구성되어 있으며 각 행동단위는 연관된 RM과 내부상태 모듈 그리고 다른 행동단위의 값을 받아서 자신의 값을 갱신한다. 또한 각 행동단위는 기억 모듈에서 모듈의 영향을 받아 학습된 내용을 적용할 수 있다[6]. 개별적인 행동단위의 구조는 그림 3과 같다.

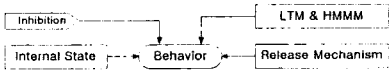


그림 3 행동단위의 구조

각 행동단위의 값을 결정하는 식은 다음과 같다[7].

$$V_{it}=[\{(1-f_{it}\times Combine((\sum_{k}RM_{kt}),(\sum_{k}IS_{nt}))\}-\sum_{m}N_{mi}\cdot V_{mt}]$$

$V_{it}$ =시간 $t$에서 행동단위 $i$의 값
$RM_{kt}$ =시간 $t$에서 $RM$ $k$의 값
$IS_{kt}$ =시간 $t$에서 내부상태 $k$의 값
$N_{mi}$ =행동단위 $i$의 다른 행동단위 $m$에 대한 억제강도
$V_{mt}$ =시간 $t$에서 다른 행동단위 $m$의 값

결정된 행동단위의 값들은 행동결정에 가장 중요한 요소로 작용한다. 행동의 결정은 각 행동단위의 값들을 기준으로 확률적으로 선택된다.

d) 기억모듈(Memory module)

기억 모듈은 단기 기억 모듈(STM Module : Short Term Memory Module), 장기 기억 모듈(LTM : Long Term Memory Module), 절차적 기억 모듈(HMMM : Hidden Markov Model memory Module)로 구성된다. 단기 기억 모듈은 현재시간에 결정된 행동과 대상을 매 시간 저장하는 기억공간이며, 이 정보를 연합 학습(Association Learning)을 통하여 장기 기억 모듈에 저장시킨다[8]. 또한 연속적인 패턴이나 정보에 관한 절차적 정보는 HMM을 통하여 동적으로 절차적 기억 모듈에 저장된다[9]. 기억모듈의 구조는 그림 4와 같다.

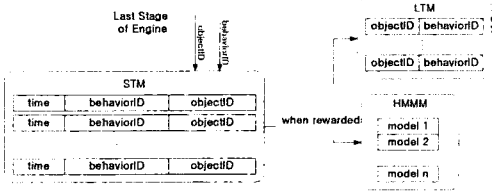


그림 4 기억모듈의 구조

e) Motor Module

행동모듈에서 결정된 행동과 object를 해석하여 외부에 전달하는 역할을 한다. 행동 선택 엔진의 출력단에 해당한다.

## 2.2 가상환경 (Virtual Environment)

행동선택 엔진은 자극에 대한 반응을 결정할 뿐 실제적인 동작을 하지는 않는다. 본 논문에서 구현한 행동선택 엔진을 적용하기 위해서는 엔진에서 나온 행동에 대한 자극의 변화를 결정하고, 변화된 자극을 다시 엔진에 전달해 줄 수 있는 환경이 필요하다. 본 논문에서는 가상 환경상의 로봇에 행동선택 엔진을 적용하였다. 가상 환경은 크게 Object들간의 상호작용을 결정하는 Object Manager부분과, 그 결과를 화면상에 출력해 주는 Display module로 나눌 수 있다(그림 1).

# 3. 행동선택 과정 및 연합학습

## 3.1 행동선택 과정

기본적인 행동선택은 외부자극의 값(RM)과 내부 상태 값(Internal State)에 의해 각 행동단위의 값이 결정되고 나면 각 행동단위들을 경쟁선택 하여 결정된다. 학습된 내용은 개별 행동단위의 값에 직접 영향을 주게 된다.

학습된 내용이 영향을 주는 과정은 다음과 같다.

센서모듈에서 전해진 자극이 메모리 모듈에 전달되고 메모리 모듈에서는 단기 기억 모듈과 절차적 기억 모듈을 검색하여 입력된 자극에 해당하는 행동-자극의 쌍이 존재하는지, 또한 그 값이 일정수준 이상인지를 판단하여, 해당하는 행동단위의 값을 증가시키게 된다. 이렇게 결정된 행동과 자극은 매 시간마다 단기 기억 모듈에 저장되며 이는 학습이 될 경우에 사용된다.

단기 기억 모듈과 절차적 기억 모듈이 갱신되는 시점즉, 학습이 일어나는 시점은 각각의 내부상태에 연결된 행동이 수행되는 경우이다. 이러한 행동이 수행되면 그 결과로 연결된 내부상태의 값을 낮춰주게 되며 기억모듈은 내부상태의 값이 일정비율 감소하면 보상을 받은 것으로 판단하여 연합 학습과 시계열 학습을 수행하게 된다

그림 5는 행동 선택과정을 나타내고 있다.

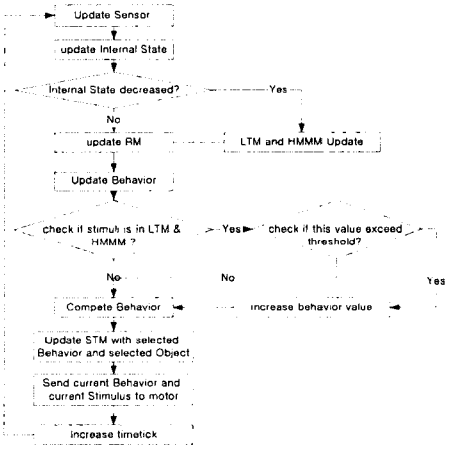


그림 5 행동선택과정 흐름도

## 3.2 연합학습 구조

연합학습(Association Learning)이란 특정 자극과 행동간의 관계를 학습하는 것으로, 고전적 조건형성과 도구적 조건형성의 2가지 종류가 있다[10]. 고전적 조건형성은 두 개 이상의 자극사이의 관련성을 학습하는 것이며, 도구적 조건형성은 반응과 그 결과사이의 관계를 학습하는 것이다.

그림 6은 단기 기억 모듈의 내용이 도구적 조건형성이 되어 장기 기억 모듈에 기록되는 과정을 나타낸다.

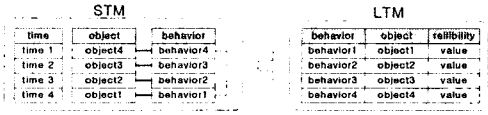


그림 6 연합학습 과정

보상을 받았을 경우 현재시간을 기준으로 일정 개수의 단기 기억 모듈의 자극과 행동이 장기 기억 모듈에 추가된다. 이미 장기 기억 모듈에 추가된 자극과 행동은 그 신뢰도(Reliability)를 아래의 식에 의해서 갱신한다.

$$R_t = R_{t-1}\times\eta(1-R_{t-1})\times d$$

$R_t$ : 자극과 행동간의 신뢰도
$\eta$ : 학습비율
$d$ : 현재시간과 자극-행동이 일어난 시간과의 차이

### 3.3 연합 학습 결과

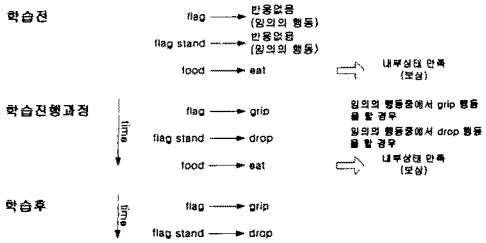


그림 7 연합 학습 과정

학습이 진행되는 도중에는 flag에 대해서 grip를 하고, flag_stand에 대해서 drop를 하는 경우에만 먹이를 줌으로서 보상을 준다(그림 7). 이를 통해 깃발을 잡고, 사용자가 원하는 장소에 가져다 놓는 일련의 자극-행동을 학습 할 수 있게 된다.

그림 8은 보상이 이루어진 시간에서의 자극과 행동을 학습한 결과이다. 그림 7은 보상이 이루어진 시간 이전의 자극과 행동을 학습한 결과이다. 보상받은 시점의 행동-자극의 관계를 학습하는 것 뿐 아니라, 지연된 보상에 대한 행동-자극의 관계도 학습할 수 있음을 알 수 있다.

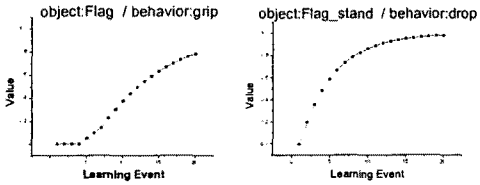


그림 7 flag-grip간의 학습곡선 그림 8 flat_stand-drop간의 학습곡선

## 4. HMM을 이용한 시계열학습

### 4.1 시계열학습 구조

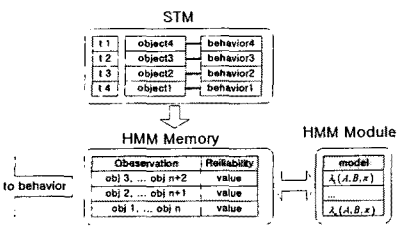


그림 9 절차적 기억모듈 구조

그림 8은 절차적 기억 모듈의 구조 및 데이터 흐름을 나타낸다. 이 모듈의 입력은 단기 기억 모듈의 연속적인 관측 정보이다. 출력은 현재 들어오는 연속적인 관측 정보를 HMM 모듈에서 학습된 모델과 비교하여 들어온 연속적 관측정보에 대한 예측 행동을 행동 모듈에 전달한다.

HMM 기억모듈로 들어온 연속적인 정보는 절차적 기억 모듈에 저장되며 그 신뢰도(Reliability)는 보상을 받을 경우 갱신된다. 절차적 기억 모듈의 신뢰도가 어느 값 이상이 되면 그 정보는 HMM 모듈에 전달된다. HMM 모듈에서는 전달된 연속 정보를 기반으로 HMM 학습을 하여 동적으로 지식을 추가하게 된다. 비교해야 할 연속 정보가 들어오면 HMM 모듈은 가지고 있는 지식, 즉 여러 HMM이 들어있는 백터와 HMM의 전향 알고리즘(Forward Algorithm)을 통하여 가장 확률이 높은 모델을 선택하게 된다. 선택된 모델을 기반으로 하여 하나의 행동을 예측하고 그 값을 행동단위에 영향을 주게 된다.

### 4.1 시계열학습 결과

시계열 학습을 통해 행동 선택 엔진은 시간에 따라 들어오는 연속적인 객체에 대한 자극을 예측할 수 있다. 행동 선택 엔진에게 입력되는 관측열은 날씨에 관한 정보로서 dark, rain, lightning이 있다. lightning 자극이 들어올 경우 goto home행동을 함으로써 두려움이 감소하여 내부상태를 만족시킬 수 있다. dark와 rain 그리고 lightning자극이 연속적으로 들어오게 되면 행동 선택 엔진은 자극의 순서를 예측하여 dark와 rain다음에 lightning이 올 것을 예측하고 dark와 rain자극이 연속해 들어올 경우 미리 goto home행동을 하여 위험을 피할 수 있다(그림 9).

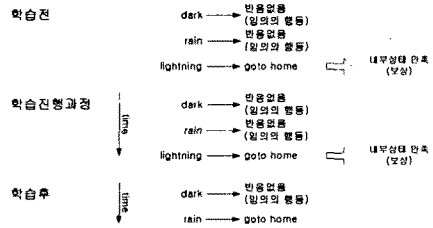


그림 10 시계열 학습 과정

## 5. 결 론

본 논문에서는 자율적으로 주어진 목적을 수행하는 지능형 소프트웨어 로봇의 행동학습 구조를 제안하고 이를 가상환경상의 로봇에 구현, 실험하여 학습결과를 검증하였다. 특히, 본 연구에서는 시간에 따른 연속적인 자극을 예측하고 적절한 행동을 선택할 수 있는 HMM 학습 모듈을 추가하였다. HMM Memory를 동적으로 관리하는 구조를 설계하여, 연속적인 시계열 패턴을 저장할 수 있게 하였고, 기억된 내용을 절차적 HMM Module과 비교하여 학습된 내용을 적용할 수 있도록 하였다. 이를 통하여 단순 자극에 대한 반응뿐 아니라 시간적 복잡성을 가진 자극에 대하여 적절한 행동을 할 수 있는 지능형 소프트웨어 로봇을 구현할 수 있었다.

본 논문에서 제안한 시계열 학습 모듈은 정해진 수의 연속된 자극에 대해서 동작하도록 구현되어 있다. 실제 환경에서는 연속된 자극의 수가 일정하지 않은 경우가 많기 때문에 이 경우에 대해서도 학습할 수 있는 구조의 설계가 요구된다. 또한 시간적 복잡성을 가진 자극뿐 아니라 공간적 복잡성을 가진 자극에도 반응할 수 있는 구조가 요구되며, 더 다양한 임무 수행을 할 수 있도록 연속된 자극에 대해 연속된 행동을 할 수 있는 구조를 설계, 구현해야 할 것이다.

### [참 고 문 헌]


[1] Maes, P., "Situated Agents Can Have Goals", Journal of Robotics and Autonomous Systems6(1&2), 1990

[2] James W.Kalat, "생물 심리학", 시그마 프레스 pp. 449, 1999

[3] David McFarLand, "Intelligent Behavior in Animals and Robots" The MIT pp. 287, 1993

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

[5] Bruce m Blumberg, " Old Tricks, NewDogs", Ethology and Interactive Createures, ,1997

[6] Toby Tyrrell "Computational Mechanisms for Action Selection" , Ph.D. Thesis, Centre for Congitive Science, University of Edinburg

[7] Bruce M. Blumberg, "No Bad Dogs: Ethological Lessons for Learning in Hmasterdam", ,1996

[8] F. von Frisch, " Honeybees: Do they use direction and distance provided by dancers?" , Science, 158 pp.1073-1076 ,1967

[9] 오영환, "음성언어정보처리" , 홍릉과학출판사 pp.52-72 , 1998

[10] R.A Rescorla, "Behavioral studies of pavlovian conditioning" , Annual review of Neuroscience, 11 pp. 329-352 ,1988