

# 인공생명체를 위한 행동선택 구조

김민조\*, 권우영\*, 이상훈\*\*, 서일홍\*

\*한양대학교 정보통신대학원 정보통신공학과

(Tel : +82-2-2290-0392;E-mail:ihshuh@hanyang.ac.kr)

\*\*한양대학교 전자 전기 제어계측공학과(Tel : +82-31-408-5802)

## Action Selection Mechanism for Artificial Life System

Min Jo Kim\* ,Woo-Young Kwon\* , Sanghoon Lee\*\*, Il Hong Suh\*

\*The Graduate school of Information & Communications , Hanyang University

(Tel : +82-2-2290-0392;E-mail:ihshuh@hanyang.ac.kr)

\*\*Dept. of EECIE., Hanyang University

(Tel : +82-31-408-5802)

### Abstract

For action selection as well as learning, simple associations between stimulus and response have been employed in most of literatures. But, for successful task accomplishment, it is required that artificial life system can learn and express behavioral sequences.

In this paper, we propose a novel action-selection-mechanism to deal with behavioral sequences. For this, we define *behavioral motivation* as a primitive node for action selection, and then hierarchically construct a tree with *behavioral motivations*. The vertical path of the tree represents behavioral sequences. Here, such a tree for our proposed ASM can be newly generated and/or updated, whenever a new behavioral sequence is learned.

To show the validity of our proposed ASM, three 2-D grid world simulations will be illustrated.

### I. 서론

인공생명체는 생명체를 모방한 가상의 시스템이다. 이러한 인공생명체는 소프트웨어 에이전트, 가상 캐릭터, 엔터테인먼트 로봇, 게임 등의 분야에 적용될 수 있다. 기존 인공지능 방법들이 이들 분야에 적용되어

왔으나 기존 인공지능 방법들은 환경에 대한 정확한 모델을 가져야 한다. 따라서 환경에 대한 불완전한 정보만을 얻을 수 있는 경우나, 변화하는 환경하에서는 적절히 동작하지 않는 문제점이 있다.

인공생명체가 지능적이며 살아있는 것처럼 보이게 하기 위해서는 주어진 환경에 적절한 행동을 선택할 수 있고, 행동에 내적인 동기나 감정을 표현할 수 있어야 한다. 또한, 과거 경험을 기반으로 환경에 적응해야 한다[2]. 이를 위해 외부 환경과, 내부 상태를 종합하여 가장 적절한 일련의 행동을 선택할 수 있는 행동선택 구조와, 변화하는 환경에 적응할 수 있는 학습 방법이 필요하다[3]. 브룩스(R. Brooks)는 이동로봇의 행동을 결정하는 구조로서 포섭구조(Subsumption Architecture)를 제안하여 인공생명체를 위한 행동선택 방법에 새로운 관점을 제시했다[4]. 이는 단순히 상태-행동에 대한 규칙들을 조합하여 복잡한 행동을 나타낼 수 있다. 또한 로렌츠(K. Lorenz), 틴버겐(N. Tinbergen)등의 동물 행동학자들은 개별적인 동물 행동의 분석을 통해 동물이 행동을 결정하는 구조를 모델화 하였다. 최근 동물의 행동선택 방식 모델을 소프트웨어로 구현하여, 로봇이나 가상 캐릭터에 적용하려는 시도가 활발히 이루어져 왔다[5][6].

그러나 기존 접근 방법은 두 가지 측면에서 한계점을 가지고 있다. 첫째로 임무수행을 위한 행동의 순서를 표현하기 어렵다. 기존 행동선택 구조는 매 순간의 상황에서 적합한 단순 행동만을 선택한다. 임무수행을 위한 행동은 사전행동들과 직접적으로 보상을 받을 수

있는 완료행동이 시간에 따라 연속적으로 나타나야만 한다. 따라서 행동의 순서를 나타낼 수 있는 행동선택 구조가 요구된다. 두 번째로 기존 행동선택 구조는 환경에 적용하는 학습 능력이 부족하다. 기존 방법은 각각의 상황에 대해 적합한 단순행동을 학습하고 이를 행동선택 구조에 추가한다. 따라서 임무수행을 위한 행동의 순서가 학습되지 못한다.

본 연구에서는 터버젠의 행동선택 모델을 수정한 행동선택 구조를 설계하고 강화학습을 통해 행동선택 구조의 구성을 갱신하는 인공 생명체에 적합한 행동선택 구조 및 학습 방법을 제안한다. 이를 위하여 2장에서 터버젠의 행동선택 모델을 설명하고, 3장에서는 행동동기를 노드로 한 행동선택 구조를 설명한다. 4장에서 제안하는 구조와 방법의 효율성을 가상 환경에서 모의 실험을 통하여 확인한다.

## II. 관련연구

동물의 행동선택에 대한 이론은 주로 동물 행동학의 분야에서 실제 동물의 행동에 대한 관찰결과를 근거로 연구되어 왔다. 로렌츠는 행동의 결정에 영향을 미치는 요소를 내부 상태와 외부 자극의 조합에 있다고 보고, 그 영향이 일정수준 이상이 되면 행동으로 표현된다고 보았다. 그러나 터버젠은 행동의 계층구조를 1차 행동동기와 2차 행동동기가 연결되어 계속 하위의 행동동기로 이어지고, 최종적으로 행동의 최소단위(예를 들어 근섬유의 운동신경)까지 도달한다고 보았다. 1차 행동동기는 주로 내부요인에 의해 영향을 받으며, 각각의 하위 행동동기들은 외부자극과 내부자극에 영향을 받는다. 하위 행동동기로의 에너지 유입(값 전달)은 하나의 블록으로 차단되어 있으며 적합한 자극이 들어오면 이 블록이 제거된다. 이러한 영향으로 가장 강하게 활성화된 본능 중심이 선택될 수 있게 한다.

동물 행동에 기반을 둔 행동선택 구조는 인공생명 연구자, 로봇 공학자들에 의해 응용되고 있다[2][7][8].

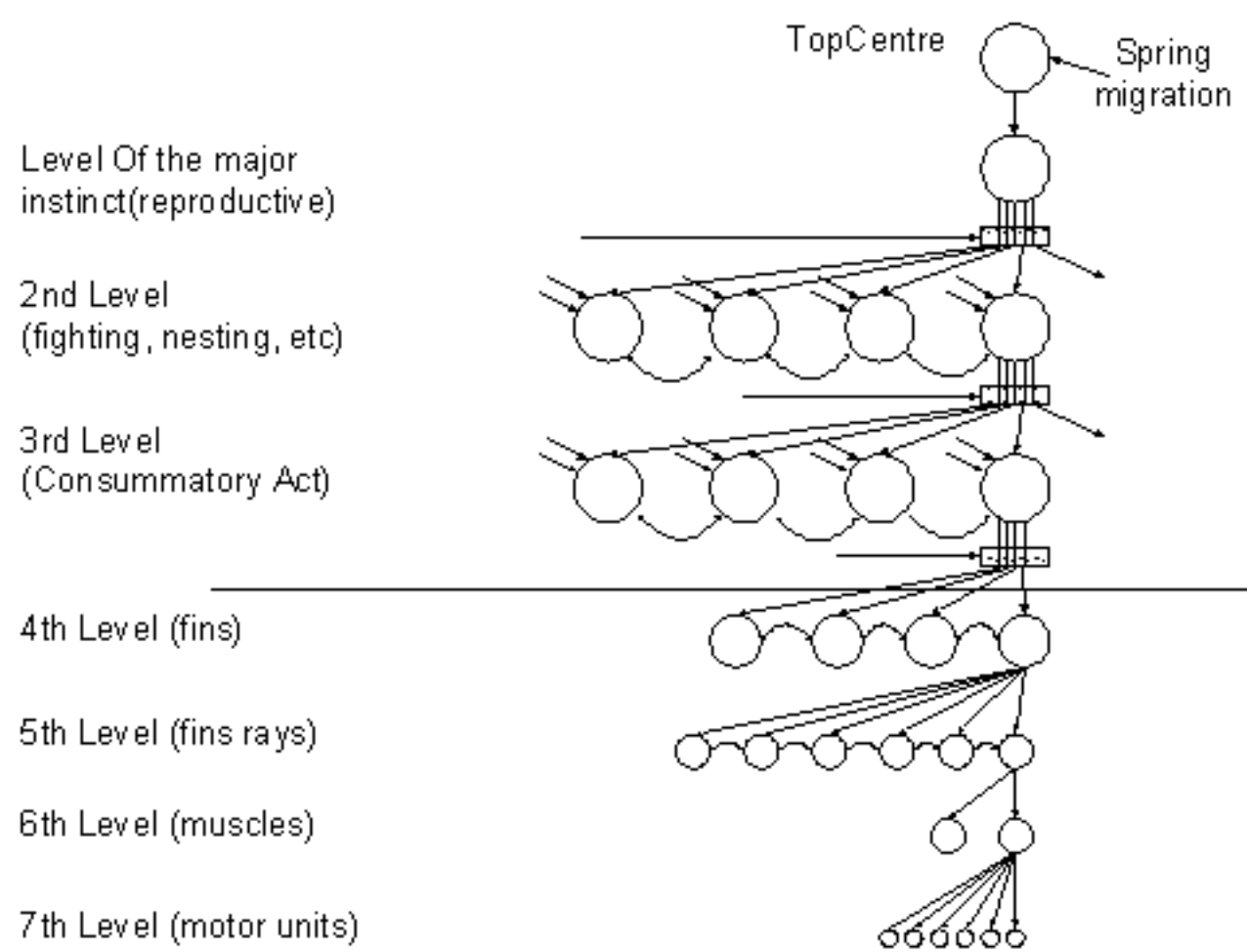


그림 1. 터버젠의 계층구조

기존 접근 방법은 여러 가지 본능이나 감정에 따라 행동을 선택할 수는 있지만, 행동의 순서를 표현하지 못하는 단점을 갖는다. 행동의 순서를 표현하기 위해서는 사전행동들과 직접적으로 보상을 받을 수 있는 완료행동을 구분하여야 하며, 이를 시간 순서에 따라 출력해야만 한다. 또한 기존 모델들은 환경에 대한 사전지식이 부족한 경우에 학습하는 능력이 부족하다.

## III. 행동선택 구조

### 3.1 행동동기

행동동기는 행동선택의 기본 단위로서 인공생명체의 내부상태와 외부 자극의 영향을 조합하여, 행동을 선택하는데 기준이 되는 값을 계산한다.

행동의 값을 계산하는 행동동기의 단위노드는 그림 2와 같으며, 각각의 행동동기의 단위노드는 자극과 내부상태 그리고 행동을 연결시켜주는 역할을 한다. 각 행동동기의 단위노드에는 고정 행동 양식이라는 행동패턴이 연결되어 있으며, 행동동기가 선택되면, 연결된 고정 행동 양식이 외부에 출력된다.

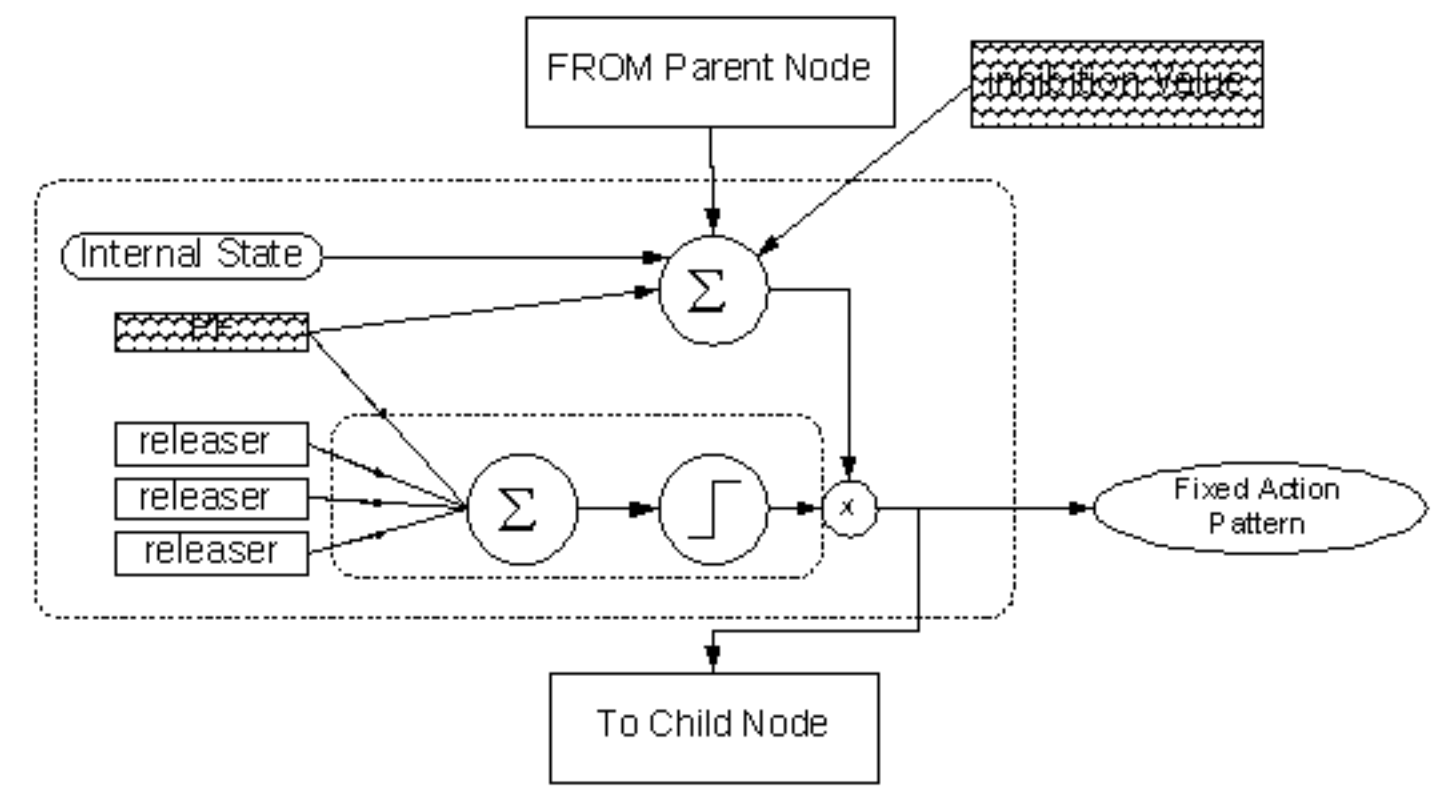


그림 2. 행동동기 기본단위의 구조

행동동기의 단위노드에 영향을 주는 요소는 다음과 같다.

- 자극인식 필터(PF : Perception Filter)  
센서를 통해 들어온 외부 환경정보 중에서 현재 행동에 필요한 자극만을 필터링 한다.

- Releaser  
행동동기를 계층구조로 구성할 경우, 하위 단계의 행동동기의 값이 결정되기 위해서는 상위 노드로부터 센서정보가 막히지 않고 하위 노드까지 전달되어야 한다. 그림 3-a는 각 행동동기의 단위노드에 외부 자극이 연결된 한 예를 보이고 있다.

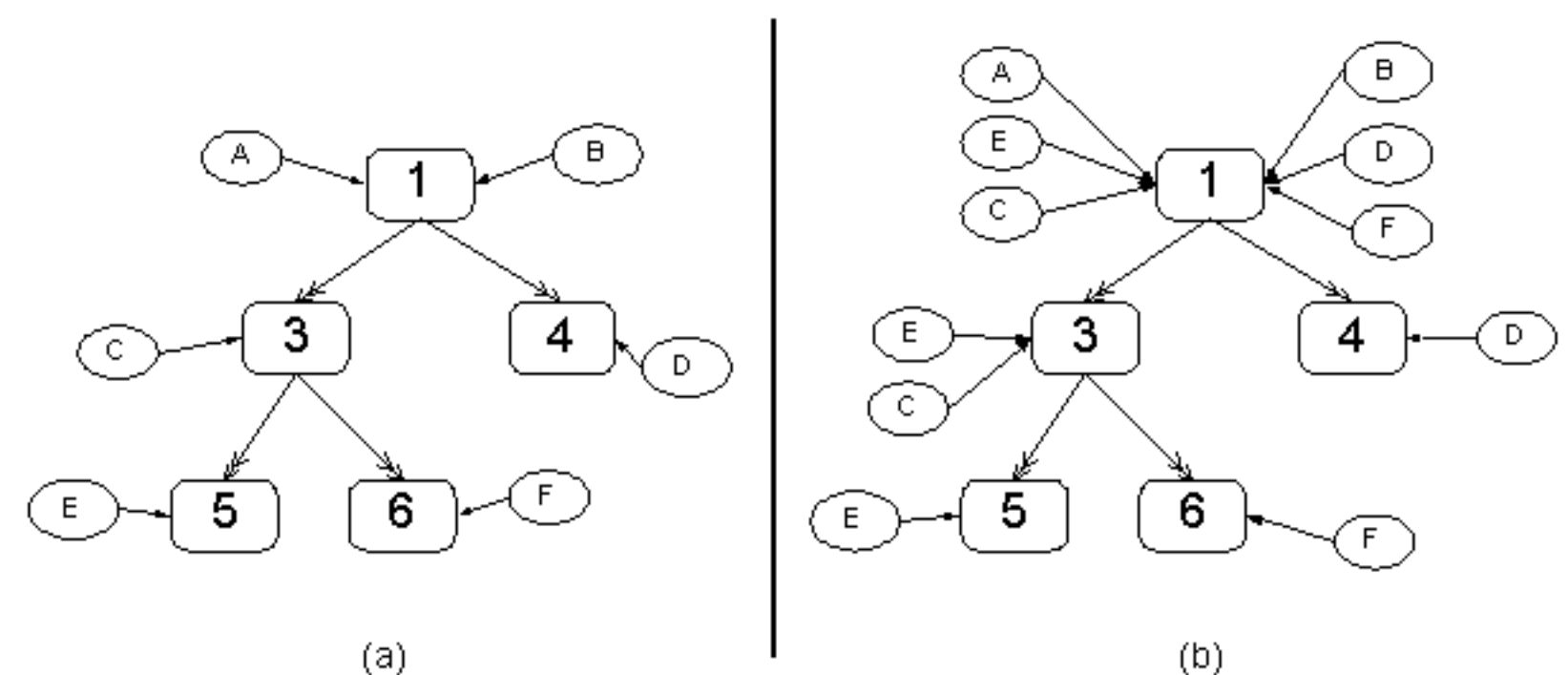


그림 3. Releaser의 사용 예

여기서 5번 노드가 선택되기 위해서는 자극 E가 들어와야 하지만, 3번 노드에서 자극 E를 통과시키지 않기 때문에 5번 노드가 선택되지 않는다. 이런 현상을 sensory bottleneck 라고 하며, 이를 해결하기 위해서는 그림 3-b에서처럼 하위 노드까지 자극이 전달될 수 있도록 상위노드에서 하위노드에 필요한 자극을 통과시켜야 한다. 1번 노드에 연결된 E, C, D, F등은 이를 위해 필요한 자극이다. 이러한 외부자극들을 releaser로 정의한다.

- 내부상태(IS : Internal State)

내부 상태는 행동에 영향을 주는 내적인 요소를 나타낸다. 즉, 욕구나 감정상태를 나타내며, 학습에 있어서 주요한 요소로 작용한다. 내부상태의 값은 최종 선택된 고정 행동 양식에 따라 변한다. 내부상태를 갱신하는 수식에서 gain은 각각의 고정 행동 양식이 내부상태에 영향을 주는 강도이며(그림 4), actionResult는 외부 상태에 의해 변하는 값이다. 한 예로 먹는 행위가 성공했을 경우 actionResult는 +1의 값을 갖게 되고 gain만큼 hunger를 감소시키게 된다. 실패했을 경우 -1의 값을 갖게 되어 hunger 및 다른 IS를 조절하게 된다. 내부상태는 식(1)에 의해 계산된다.

$$IS_i = IS_i + \Delta IS_i$$

$$\Delta IS_i = actionResult_j \times gain_{ij} \quad (1)$$

i : index of Internal State

j : index of selected Fixed Action Pattern

Action / Variables	Thirst	Hunger	Fatigue	Sleep
Move	+1	+1	+4	+3
Drink	-A / +1	+1	+1	+3
Eat	+1	-A / +1	+1	+3
Sleep	+1	+1	-15	-20
Nothing	+1	+1	+1	+3

그림 4. 고정 행동 양식과 내부상태와의 관계

- 상위 행동동기

행동동기들은 최상위 단계의 행동동기에서 시작하여 그 값을 하위로 전달하는 계층구조로 이루어진다. 상위단계에 행동동기의 값을 받아 자신의 값을 갱신하며, 이를 다시 하위 노드로 전달하게 된다.

위와 같은 요소에 의해 각 행동동기의 값이 결정되며 행동동기의 값이 갱신되는 식은 식(2)와 같다.

$$V_{BM_i} = \left( parentV_{BM} + \sum V_{V_i} + \sum V_{PF_i} \right) STEP \left( \sum V_{Releaser_m} \right) \quad (2)$$

$$STEP(x)$$

if x=0 STEP(x)=0  
else STEP(x)=1

3.2 행동동기의 계층구조

행동동기의 계층구조는 두 가지 목적을 위해 사용된다. 첫 번째 목적은, 인공생명체의 내적인 본능에 따라 행동의 동기들을 계층적으로 구성하여 현 상태에서 가장 적절한 행동을 할 수 있게 하기 위함이다. 또 다른 목적은 행동의 순서를 표현하기 위해서이다. 기존의 계층구조(예를 들어 터버겐의 본능모델)에서는 주어진 상태에서 하나의 완료행동만을 표현할 수 있다. 그러나 일반적인 경우 임무를 수행하기 위한 행동은 사전 행동들과 완료행동이 시간순서에 따라 나타나야 한다.

행동의 순서는 각각의 자극에 대한 반응들이 연쇄되면서 나타나게 된다. 그림 5는 음식과 물을 먹는 경우에 행동의 순서를 제안한 계층구조로 나타낸 예이다.

행동의 순서를 이루는 행동동기들은 자신에게 맞는 자극을 선택하여 행동이 선택되도록 하며, 다음 자극이 나타날 때까지 하위 노드에 값 전달을 봉쇄한다. 따라서 계층구조의 수직 경로는 행동의 순서를 나타낼 수 있다. 예를 들면 hunger라는 내부상태가 높고 food exist PF가 활성화되면 2번 행동동기가 선택되어 move 행동이 나타난다. 이 경우 food near 및 food contact PF는 비활성화 되어 있기 때문에 하위 노드인 3번 노드에 값 전달은 막혀있게 된다. 시간이 지나고 food에 가까이 가게 되면 food exist PF와 food near PF가 동시에 활성화되어 3번 노드가 선택되고 get이라는 행동이 나타난다. 이런 순서로 최하위 노드까지 행동을 반복하면 전체 행동이 완료되고 보상을 받을 수 있게 된다.

계층구조의 첫 번째 단계는 내부상태의 영향을 받게 되며 그 하위 노드에 내부상태를 만족시키기 위한 행동의 순서가 연결된다. 행동이 선택되는 과정은 첫 번째 단계의 행동동기들 중 가장 큰 값을 갖는 행동동기가 선택된다. 선택된 행동동기의 하위 행동동기들 중 가장 큰 값을 가진 행동동기가 다시 선택되며 이 과정은 하위 행동동기의 값이 모두 0일 때까지 반복된다.

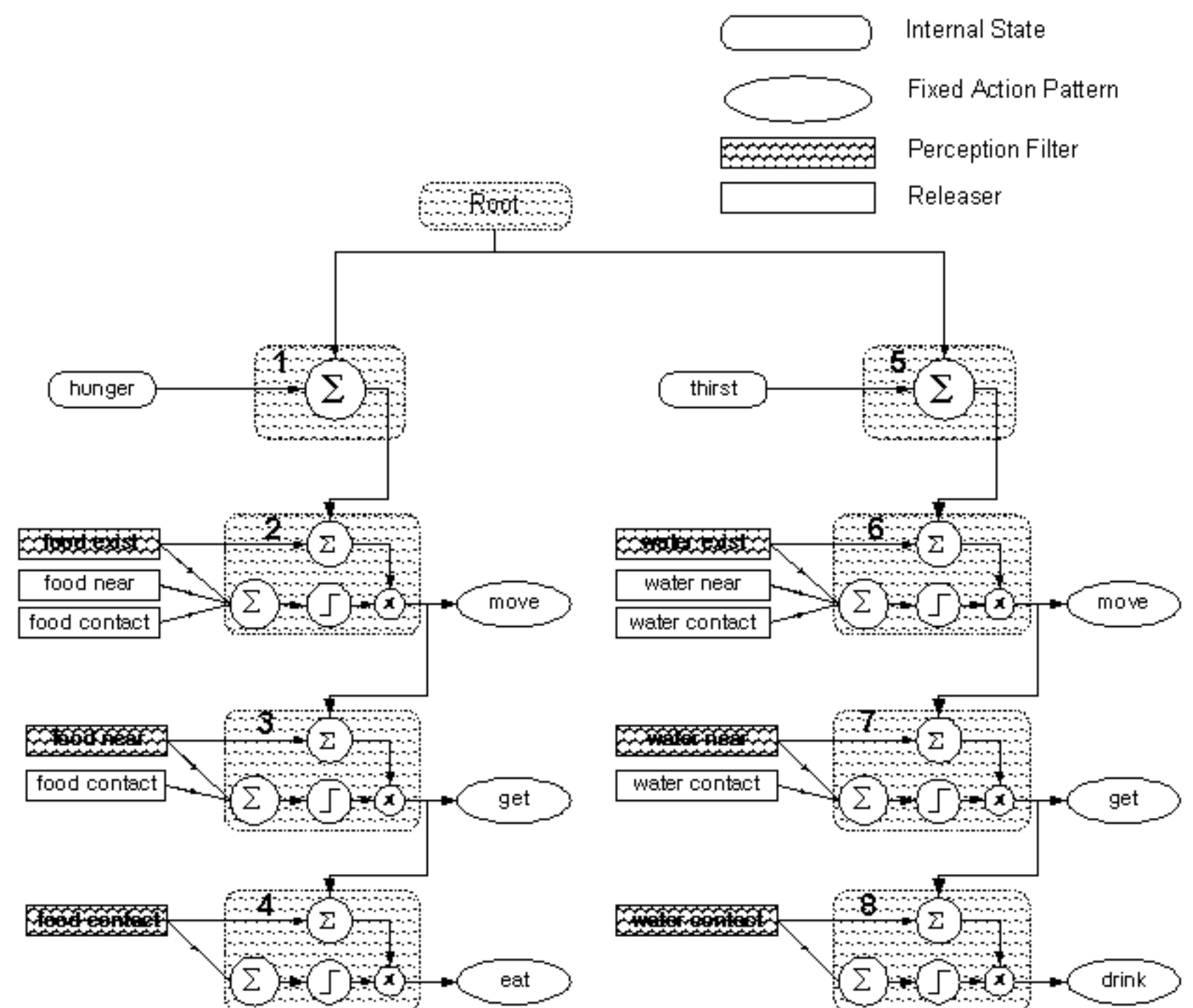


그림 5. 행동동기의 계층구조



### 3.3 학습절차

기존 행동동기의 계층구조에서 행동동기가 선택되지 못할 경우에는 탐색을 수행하게 된다. 탐색 과정은 자극과 고정 행동 양식을 임의 선택하여 이루어지며, 그 과정은 단기기억(STM : Short Term Memory)에 저장된다. 단기기억에 저장된 내용은 보상을 받을 경우 장기기억(LTM : Long Term Memory)으로 전환되며, 그 과정에서 자극과 행동 그리고 행동의 결과에 대한 신뢰도를 기록한다. 이 과정은 강화학습 방법에 의해 이루어지며(그림 6), Monte Carlo 방법과 유사하다. 신뢰도를 갱신하는 방법은 식(3)에 나타나 있다.

$$V(s_i, a_i, s_{i+1}) = V(s_i, a_i, s_{i+1}) + \frac{\eta(1 - v(s_i, a_i, s_{i+1}))}{(N - i + 1)^2} \quad (3).$$

본 행동선택 구조에서 학습은 내부상태 즉 욕구를 만족시키기 위한 행동의 순서를 배우는 과정이다. 따라서 장기기억은 내부상태 마다 각각 다른 기억공간을 가진다. 장기기억 내용의 신뢰도가 일정 수준 이상이 되면 행동동기의 계층구조에 추가되며 그 과정은 그림 7과 같다.

다음으로 그림 8은 임계값을 넘은 장기기억 요소들이 행동동기의 계층구조에 추가된 결과를 예를 들어 보이고 있다.

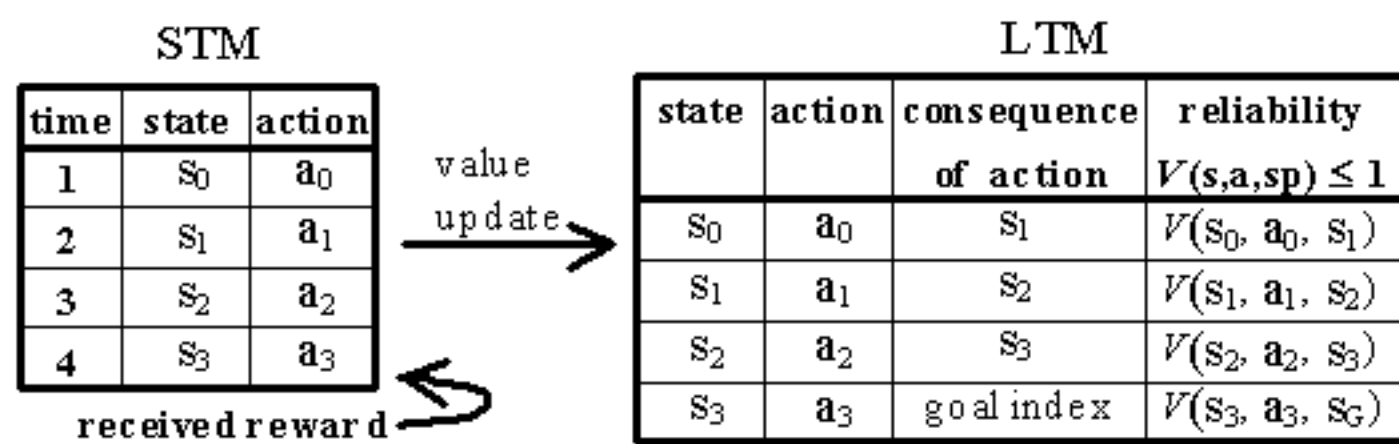


그림 6. 단기기억에서 장기기억으로 전환되는 과정

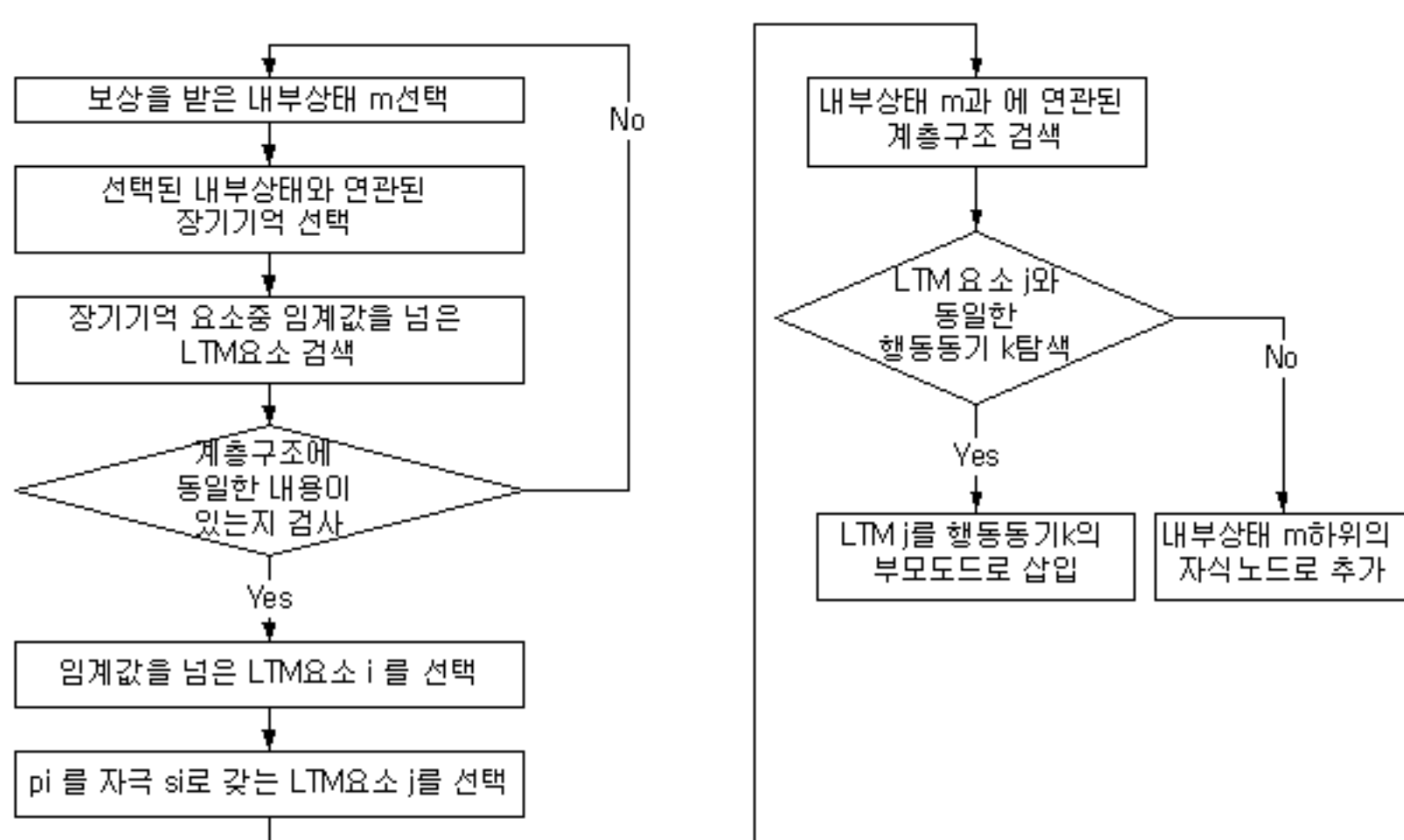


그림 7. 행동동기가 계층구조에 추가되는 과정

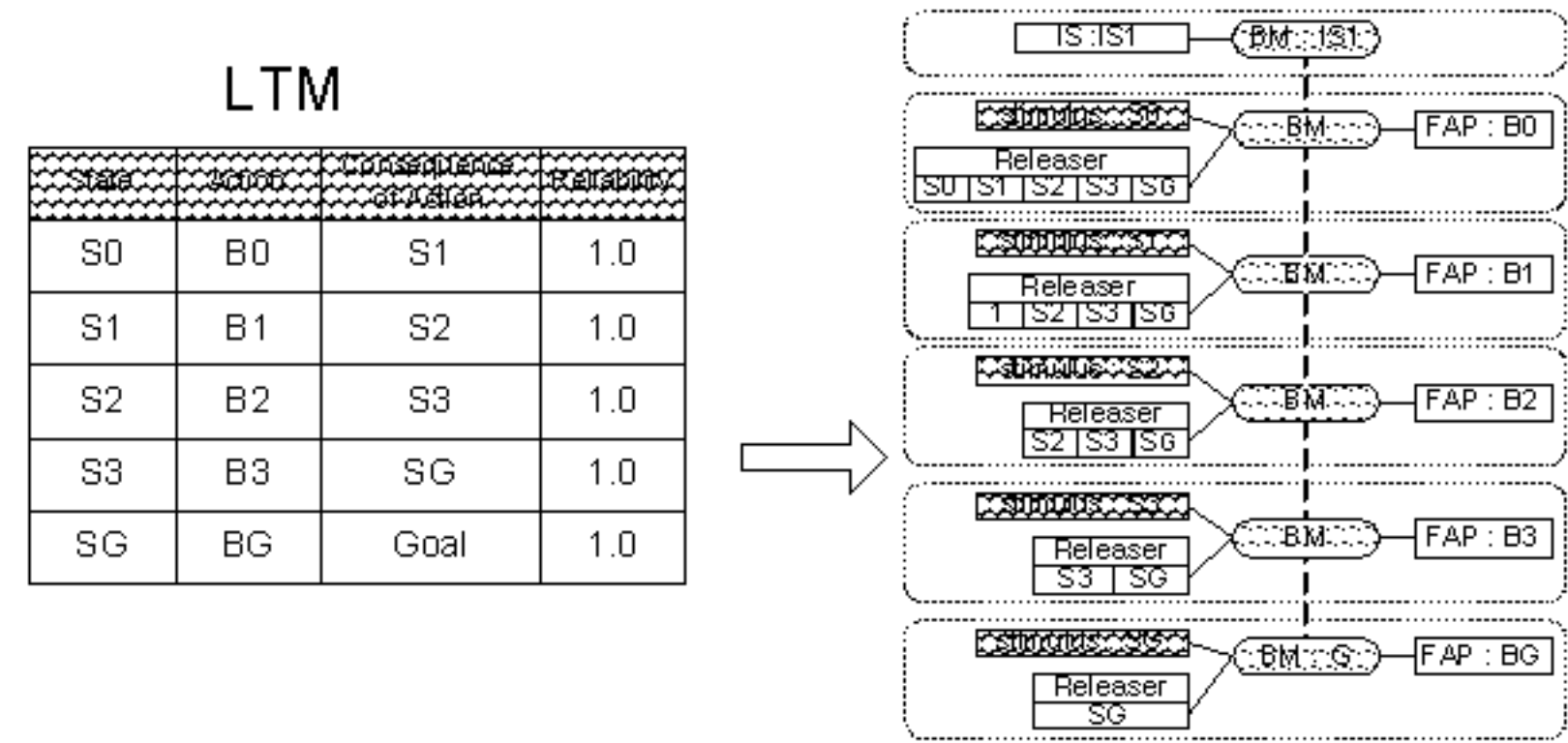


그림 8. 장기기억 요소들이 행동동기의 계층구조에 추가된 결과

## IV. 구현 및 실험

### 5.1 실험환경

제안된 행동선택 구조를 검증하기 위하여 2차원의 격자환경에서 가상 캐릭터를 구현하여 실험하였다. 가상 캐릭터는 동, 서, 남, 북 방향으로 움직이는 행동을 할 수 있으며, 4x4의 격자 내에 있는 모든 물체를 인식할 수 있다. 가상 환경에서는 블록과 벌레가 존재한다. 학습해야 할 규칙은 다음과 같다.

1. 캐릭터-블록-벌레가 일직선을 이룰 때 블록을 밀면 벌레를 죽여 보상을 받을 수 있다(그림 9-a).
2. 블록과 벌레가 일직선을 이룰 때 캐릭터는 블록을 밀 수 있는 상황까지 이동한다(그림 9-b).
3. 블록과 벌레가 일직선을 이루지 않을 경우 블록을 밀어서 일직선을 이루게 한다(그림 9-c).

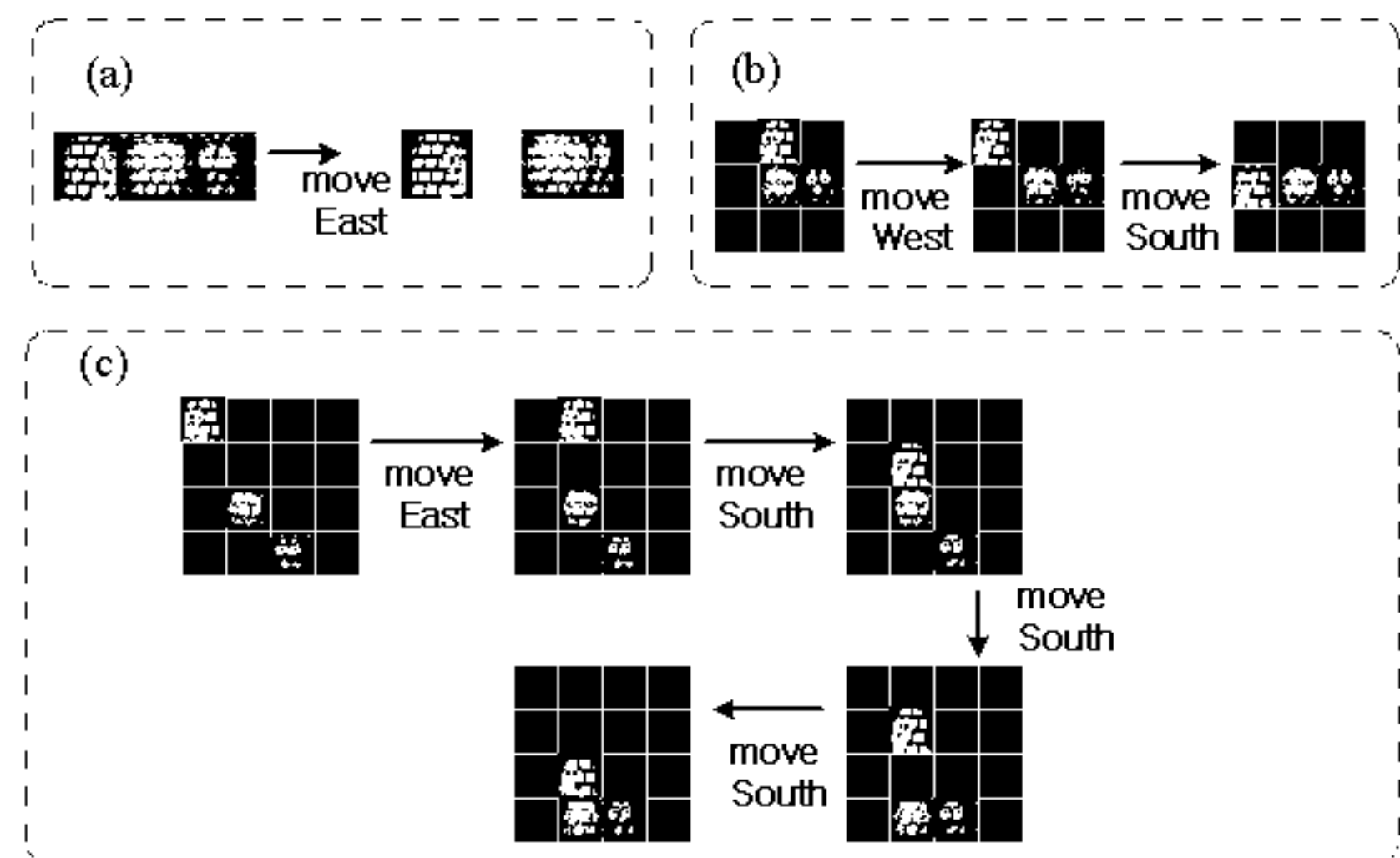


그림 9. 학습해야 할 규칙들

실험과정에서는 먼저 단순한 규칙을 학습하여 계층구조에 추가하고, 추가된 규칙을 활용하여 보다 복잡한 규칙을 학습하도록 하였다. 규칙 1을 학습하기 위하여 3x3의 격자에서 50회의 시도를 반복하였으며, 규칙 2를 학습하기 위해서 규칙 1과 동일한 환경에서 100회의 시도를 반복하였다. 마지막으로 규칙 3을 학습하기 위해서 4x4의 격자에서 1000회의 시도를 반복하여 행동동기의 계층구조에 추가할 수 있었다.

## 5.2 실험 결과

실험이 수행된 후 계층구조에는 보상을 받기 위한 행동동기들이 추가되었음을 볼 수 있었다. 그림 10과 그림 11은 각각 규칙 1을 학습한 후와 규칙 3까지 학습된 이후, 추가된 행동동기의 계층구조 일부를 나타내고 있다.

새롭게 추가된 행동 순서의 예로 그림 10에서 먹이를 섭취하려는 행동동기를 성공적으로 만족시키기 위한 사전행동이 추가되었으며, 실험에서 해당 행동의 순서에 따라 행동이 선택되었다. 한 예로 상태 1에서 move W, 상태 2에서 move W, 상태 3에서 move S, 상태 4와 5에서 move E 행동을 하는 일련의 행동 순서를 보임으로서 보상을 받을 수 있었다.

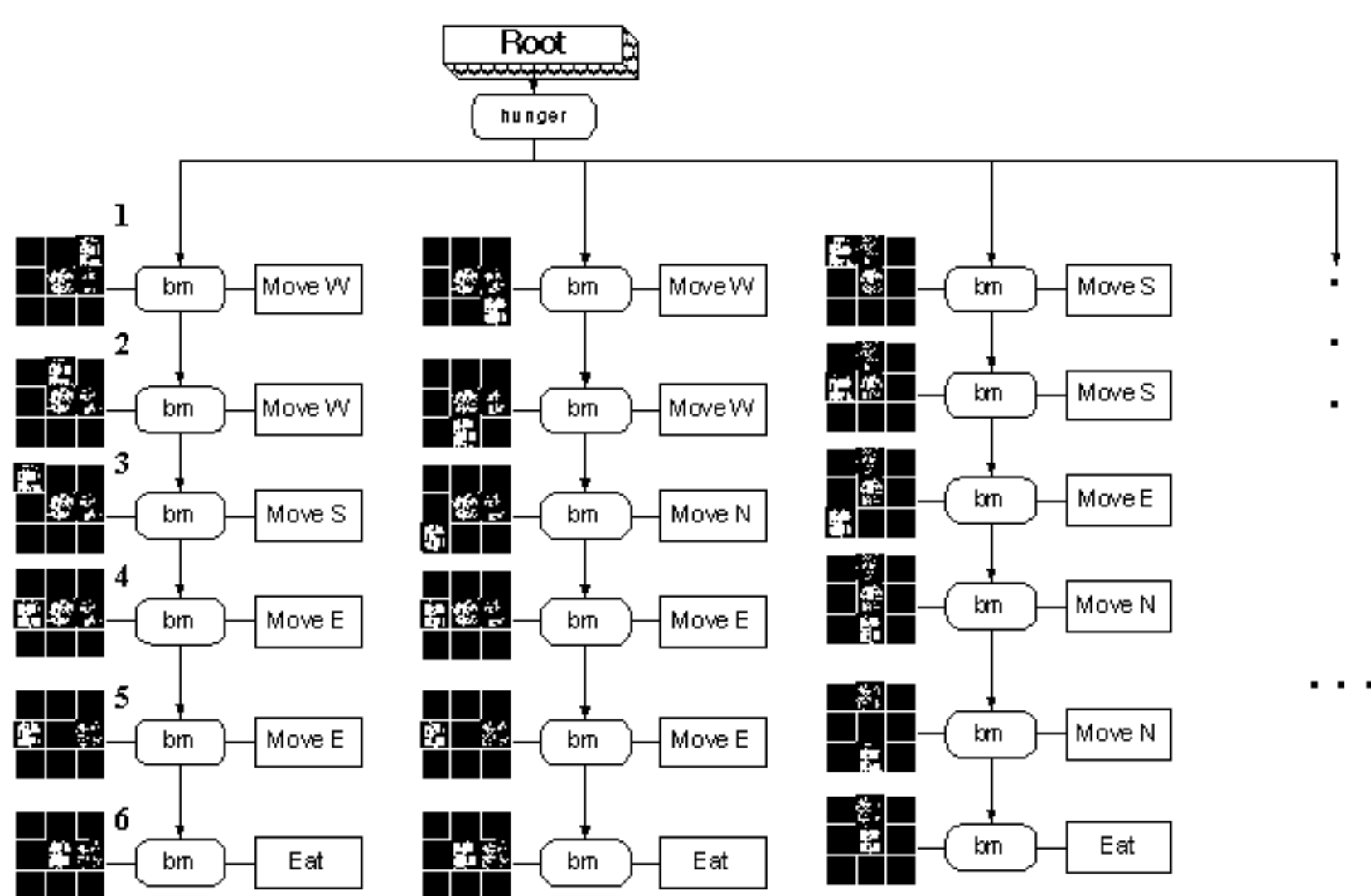


그림 10. 규칙 1,2를 학습한 이후의 행동동기의 계층구조

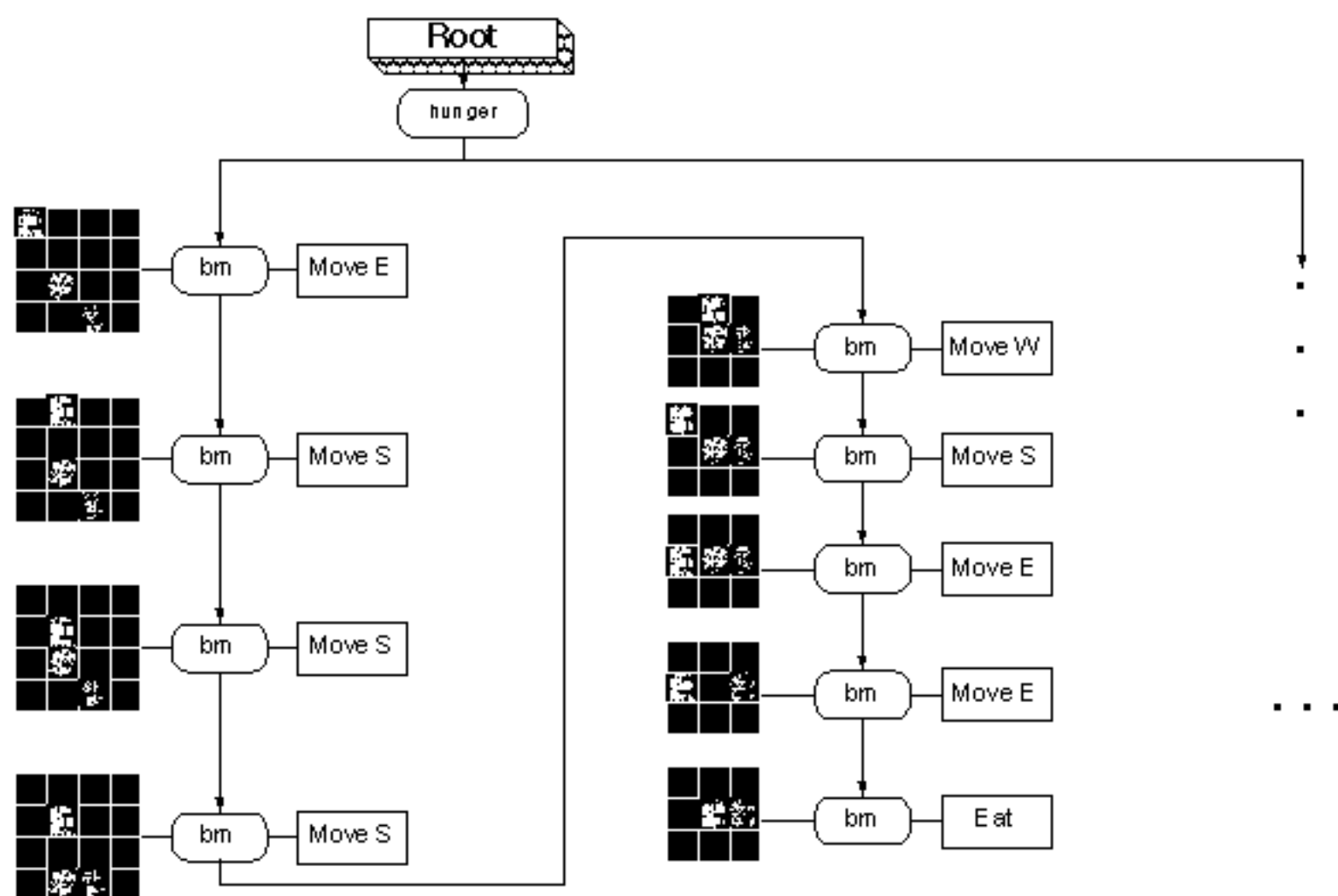


그림 11. 규칙 3을 학습한 이후의 행동동기 계층구조

## V. 결론

행동선택 구조에 대한 기존 연구결과들이 행동의 순서를 학습하고 이를 행동선택 구조에 추가하는 부분에서 문제점이 있었기 때문에, 본 연구에서는 이를 보완한 새로운 행동선택 구조를 구현하였다. 구현된 행동선택 구조에서는 외부자극과 행동 패턴간의 연결 관계를 조절하는 행동동기를 정의하여, 각각의 상황에 적

절한 행동을 할 수 있었다. 또한 행동동기들을 계층구조로 구성하고 계층구조의 수직경로가 행동의 순서를 표현하도록 하였다. 하나의 행동이 완료되기 전까지 다음 행동을 봉쇄함으로써 행동의 순서를 나타낼 수 있으며, 학습된 행동의 순서는 새로운 트리 형태로 행동선택 구조에 자동으로 추가되었다.

제안한 행동선택 구조와 학습방법의 타당성을 검증하기 위하여 2차원의 격자환경에서 실험을 하였다. 단순한 규칙을 먼저 학습하여 계층구조에 추가하고, 추가된 규칙을 활용하여 보다 복잡한 규칙을 학습하도록 하여 학습 시간을 단축하였다. 결과로서 임무를 수행하기 위한 일련의 행동 순서들이 생성되어, 행동동기의 계층구조에 추가되는 과정을 확인하였다.

## 참고문헌

- [1] Levy Steven, *Artificial Life*, Nework: Vintage Books, 1992.
- [2] Bruce M Blumberg, *"Old Tricks, NewDogs"*, Ethology and Interactive Createures, ,1997
- [3] P. Maes., *"Modeling Adaptive Autonomous Agents."*, Artificial Life Journal, edited by C. Langton, Vol. 1, No. 1 & 2, pp. 135-162, MIT Press, 1994.
- [4] R.Brooks, *"A Robust Layered Control System For a Mobile Robot."*, In IEEE Journal of Robotics and Automation, pages 14 23, April, 1986.
- [5] N.Tinbergen, *The Study of Instinct.*, Oxford University Press., 1951.
- [6] Lorenz, K. *Foundations of Ethology.* ,Springer-Verlag, New York, 1973
- [7] P. Maes. *,"A Bottom-Up Mechanism for Behavior Selection in an Artificial Creature"*, In From Animals to Animats, First International Conference on Simulation of Adaptive Behaviour, MIT Press, Cambridge, 1991
- [8] Toby Tyrrell *"Computational Mechanisms for Action Selection"* , Ph.D. Thesis, Centre for Congitive Science, University of Edinburg ,1993
- [9] R. Sutton , A. Barto, *Reinforcement Learning*, MIT Press, 1997.