

Q-learning과 Cascade SVM을 이용한 군집로봇의 행동학습 및 진화

Behavior Learning and Evolution of Swarm Robot System using Q-learning and Cascade SVM

서상욱 · 양현창 · 심귀보*

Sang-Wook Seo, Hyun-Chang Yang and Kwee-Bo Sim*

중앙대학교 전자전기공학부

요 약

군집 로봇시스템에서 개개의 로봇은 스스로 주위의 환경과 자신의 상태를 스스로 판단하여 행동하고, 필요에 따라서는 다른 로봇과 협조를 통하여 어떤 주어진 일을 수행할 수 있어야 한다. 따라서 개개의 로봇은 동적으로 변화하는 환경에 잘 적응할 수 있는 학습과 진화능력을 갖는 것이 필수적이다. 본 논문에서는 SVM을 여러 개 이용한 강화학습과 분산유전알고리즘을 이용한 새로운 자율이동로봇의 행동학습 및 진화방법을 제안한다. 또한 개개의 로봇이 통신을 통하여 염색체를 교환하는 분산유전알고리즘은 각기 다른 환경에서 학습한 우수한 염색체로부터 자신의 능력을 향상시킨다. 특히 본 논문에서는 진화의 성능을 향상시키기 위하여 Cascade SVM을 기반으로 한 강화학습의 특성을 이용한 선택 교배방법을 채택하였다.

Abstract

In swarm robot systems, each robot must behaves by itself according to the its states and environments, and if necessary, must cooperates with other robots in order to carry out a given task. Therefore it is essential that each robot has both learning and evolution ability to adapt the dynamic environments. In this paper, reinforcement learning method using many SVM based on structural risk minimization and distributed genetic algorithms is proposed for behavior learning and evolution of collective autonomous mobile robots. By distributed genetic algorithm exchanging the chromosome acquired under different environments by communication each robot can improve its behavior ability. Specially, in order to improve the performance of evolution, selective crossover using the characteristic of reinforcement learning that basis of Cascade SVM is adopted in this paper.

Key Words : Behavior Learning, Evolution, Swarm Robot, Cascade SVM, Q-learning

1. 서 론

최근 로봇 응용분야의 확장으로 군집로봇으로 구성된 시스템에 대한 연구가 많이 이루어지고 있다. 특히 자연계 생물의 특징인 자율 분산성을 가지는 군집 로봇시스템에 관한 연구가 관심을 모으고 있다. 이러한 군집 로봇시스템은 중앙관리형 시스템에 비하여 다음의 몇 가지 특징을 갖는다. 첫째로 각각의 로봇은 주변의 환경이나 물체, 다른 로봇의 행동 등을 인식하여 자신의 행동을 독립적으로 결정하며, 또한 주어진 작업을 잘 수행하기 위하여 다른 로봇과 협동할 수 있다. 둘째로 자율분산로봇시스템은 강건성(robustness)과 유연성(flexibility)을 가지고 있다. 몇 대의

로봇이 고장이 나더라도 시스템의 정상적인 동작에 영향을 주지 않으며, 주어진 일에 대하여 오직 로봇의 행동 규칙만 바꾸어 줌으로써 여러 가지 작업에 적용할 수 있다. 셋째로 시스템의 크기가 커지더라도 개개의 로봇의 자신의 주변상황에 따라 자신의 일을 판단하여 결정하므로 시스템의 복잡도가 증가하지 않는다[1][2].

다수의 자율이동로봇에 의한 효과는 크기는 군(swarm) 전체의 거동을 나타내는 군 행동과 작게는 시스템내의 작업을 수행하는 협조 작업으로서 나타나게 된다. 일반적으로 군 전체의 이동이나 배치 등과 같은 군 행동은 센싱 능력에 의해서 충분히 구현할 수 있다. Mataric은 통신의 기능은 사용하지 않고 센서만을 사용해 주위 환경을 인식하고 행동을 취하는 시스템에 대하여 연구하였으며 이에 의해 여러 가지 군 행동을 실현하였다[3]. 그러나 협조 작업의 경우 통신에 의해 자신의 상태 및 정보를 교환함으로써 쉽게 협조 작업을 수행할 수가 있다. 통신기능을 사용하지 않음으로써 나타나는 결과들은 전체적인 집단행동 등의 범위로 한정되었으며 구체적인 팀의 구성이나 특정한 작업을 하기 위한 협조는 고려되지 않았다.

군집 로봇시스템에서 개개의 로봇은 사실상 다른 모든

접수일자 : 2008년 11월 1일

완료일자 : 2009년 1월 31일

* 교신 저자

본 연구는 지식경제부의 2008년도 성장동력기술개발사업인 「집단 로봇 기술을 이용한 사회안전로봇 개발(세부과제: 로봇통제 및 환경기술개발)」에 의해 수행되었습니다. 연구비 지원에 감사드립니다.

로봇의 정보를 알 필요가 없으며 자신이 처한 환경만 인식하여 행동하면 된다. 그러나 일반적으로 동적으로 환경이 변화하는 시스템에서 로봇 스스로 협조를 위한 최선의 행동을 결정하는 것은 매우 어렵다. 최근 자연계의 생물체의 구조 및 거동을 인공적으로 연구하는 인공생명의 방법이 이와 같은 예측이 불가능하고 복잡한 문제를 해결하는데 새로운 해결책으로 기대되고 있다[4-5].

인공생명 기법에 의한 로봇은 기존의 지능로봇에 더하여 다음과 같은 세 가지의 특징을 갖는다. 첫째로 문제의 수행에 있어서 사전에 짜여진 완벽한 계획보다는 예측하지 못한 문제가 발생하였을 경우 즉각적인 대처와 참여에 의해 적응 및 학습을 해나가는 능력이 있다. 따라서 행동계획은 보다 자연스럽고 유동적이게 하여 환경조건으로부터 발현될 수 있도록 한다. 이를 위하여 로봇 설계자는 완벽한 사전계획보다는 로봇 스스로가 문제를 해결할 수 있는 구조를 만들어주는 것이 필요하다. 두 번째로 개체 간 또는 환경과의 상호작용에 의해 창발적인 행동이 나타난다. 세 번째로 자연계에서 개체는 다른 개체에 대한 관찰과 모방(흉내)을 통하여 학습을 한다. 이것은 개체가 부가적인 지식을 얻는 매우 실제적인 방법으로 로봇에게 적용하면 로봇이 자신의 프로그램을 학습하는데 좋은 결과를 얻을 수 있는 바탕이 될 수 있을 것이다. 그러나 대부분의 로봇은 다른 로봇이 무엇을 하는지, 또 어디로 가는지에 대한 것을 감지할 수 있는 충분히 발전된 인지능력을 가지고 있지 못하다. 이러한 한계점 때문에 관찰과 모방에 의한 학습을 실현하기 어렵다.

본 논문은 자율분산로봇시스템에서 자율적으로 행동하며 시스템의 목적을 달성하는 로봇을 실현하기 위하여 사전에 짜여진 완벽한 계획이 아닌 시스템에 적용할 수 있는 구조를 설계하여 주었다. 로봇은 주어진 환경에서 자신의 행동을 학습하기 위하여 Cascade SVM을 기반으로 한 강화학습을 이용하였고 진화를 위하여 분산유전알고리즘을 도입하였다. 기본적으로 강화학습은 환경에 대한 사전지식이 없는 경우 강화신호에 의하여 행동을 학습시켜 나가는 방법을 말하는데, 여기서 제안한 Cascade SVM을 기반으로 한 강화 학습은 즉각적인 보상 대신 Pattern classifier를 사용하여 로봇의 행동에 대한 옳고 그름에 관한 분류를 나눈 후 옳은 행동에 대하여 보상을 주는 방식을 선택하였다. 로봇의 행동은 일반적으로 비선형이므로 입력 데이터 값을 커널 함수를 이용해 보다 높은 특징 공간으로 맵핑하여 보다 높은 차원의 유클리드 공간으로 학습데이터를 투영하였다[6].

본 논문은 2장에서 군집 로봇의 통신 모델 중 융합 모델에 관한 설명을 하였고, 3장에서는 Cascade SVM을 기반으로 한 Q-학습방법을 사용한 로봇의 행동학습에 대한 설명을 하였으며, 4장에서는 분산유전알고리즘에 의한 행동진화에 관하여 설명을 하였다. 그리고 5장에서는 자율이동로봇 시스템의 행동학습과 진화를 위하여 다수 로봇에 의한 물체 획득 문제로 설정한 시뮬레이션의 결과에 대해서 설명하였고, 마지막으로 6장에서는 결론에 대해서 설명하였다.

2. 군집 로봇의 통신 모델

로봇간의 통신은 통신 범위에 따라 전역적 통신과 지역적 통신으로 나눌 수 있다. 또한 수신자의 여부에 따라 특정한 수신자에게 정보를 보내는 정보전달(message passing) 모델과 어떤 특정한 수신자가 정보를 받을 것을 기대하지 않고 정보를 내보내는 사인보드(sign board) 모델, 이

두 가지 방법을 동시에 사용할 수 있도록 한 융합 모델 세 가지로 나눌 수 있다. 본 논문에서는 로봇의 지역적 통신에서 이 두 가지 모델을 융합한 통신 모델을 사용한다.

2.1 데이터 형식

융합 모델의 데이터는 그림 1과 같이 앞의 헤더(header) 부분과 데이터 끝의 테일(tail) 부분 사이에 사인보드의 내용과 정보 전달을 위한 정보가 차례로 들어가며 특정한 로봇에게 보낼 정보가 없을 경우 Message 부분은 생략된다. 즉 정보를 보낼 대상의 로봇이 주위에 있을 경우에만 모든 송신부에서 Message 부분의 정보를 추가해서 보내게 된다.

Header	Sign Board	Message	Tail
--------	------------	---------	------

그림 1. 통신 데이터 형식

Fig. 1. Communication Data Form

Sign board에 포함되는 내용은 다음과 같은 것이 있다. ①로봇의 고유 번호(ID), ②송신부의 번호, ③이동 속도, ④로봇의 상태: 고장, 작업 중, 대기 중 등, ⑤시스템 내에서 획득한 정보(information), ⑥도움 요청의 여부 등이다.

2.2 협조 행동을 위한 팀 구성

시스템 내에서 두 대 이상의 로봇이 함께 협조 작업을 해야 할 경우가 발생했을 때 통신에 의해 두 대 이상의 로봇이 팀을 이루게 된다. 작업 환경 내에 필요로 한 로봇이 한 대일 경우 그 환경 내에 있는 가장 가까운 로봇이 작업을 하게 된다. 만일 로봇이 작업 환경 내에서 임의로 움직이다 협조가 필요한 작업을 발견하면 사인보드의 형태로 주위의 로봇에게 도움을 요청하게 된다. 이때 응답한 로봇이 있으면 그 로봇은 서브(sub)가 되고 도움을 요청한 로봇은 메인(main)이 되어 팀을 구성한다.

3. 행동학습 및 진화시스템의 구조

그림 2는 자율분산로봇시스템의 행동학습 및 진화를 위한 군집 로봇시스템의 개념도이다. 각각의 로봇은 다수의 상태-행동 규칙을 테이블의 형태로 가지고 있으며, 테이블의 값은 행동 결과에 의해 주어진 보상이나 벌칙에 따라 제안한 Cascade SVM을 기초로 한 Q-학습의 방법으로 갱신해 나간다. 만일 로봇이 자신보다 우수한 로봇을 만났을 경우 이 로봇은 통신을 통해 상대방의 행동규칙을 획득하고, 분산유전알고리즘에 의해 자신의 행동규칙을 진화해 나간다.

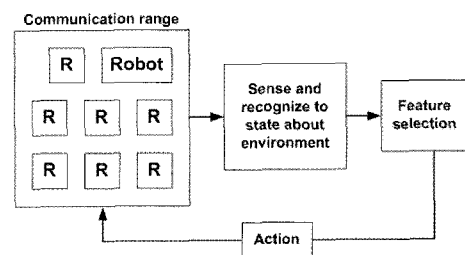


그림 2. 군집 로봇시스템의 행동학습 및 진화의 개념도
Fig. 2. Conceptual diagram of behavior learning and evolution

- Step 1. 모든 상태 s 와 행동 a 에 대하여 $Q(s, a)$ 을 임의의 값(일반적으로 0)으로 초기화한다.
- Step 2. 현재의 상태 s 를 인식한다.
- Step 3. (상태, 행동) 규칙에 따라 행동 a 를 선택한다.
- Step 4. 주어진 환경에서 행동 a 를 수행하고, 다음상태를 s' 즉각적인 보상을 r 로 놓는다.
- Step 5. s, a, s', r 로부터 (상태, 행동) 규칙을 다음의 식 (1)에 의해서 갱신한다.

$$Q_{t+1}(s, a) = (1 - \alpha_t) Q_t(s, a) + \alpha_t (r + \gamma \max_{a' \in \mathbf{A}} Q_t(s', a')) \quad (1)$$

식에서 α_t 는 학습률이고, γ 는 0과 1사이의 고정된 감쇠계수이다.

- Step 6. step2로 되돌아간다.

3.2 SVM(Support Vector Machine)

로봇의 행동 상태는 로봇에 소속된 환경에 대하여 센싱과 인식을 한 상태 값들을 SVM에 의하여 패턴을 분류하고 최적화된 상태에 따라 행동을 하게 된다. Correct와 incorrect 두 가지로 나누어 correct 상태가 나온 경우 그 행동 상태에 가중치를 두고 uncorrect가 나온 경우 행동 상태를 버리는 과정을 거친다. 그림 3은 자율분산로봇시스템의 행동 학습 부분을 구체화시킨 그림이다.

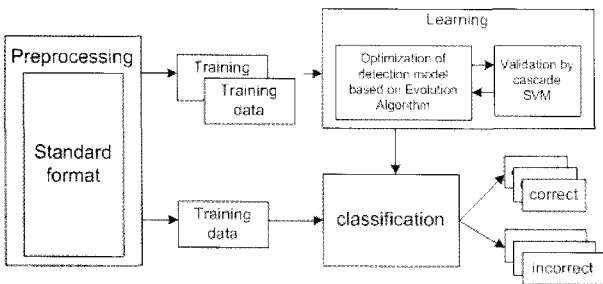


그림 3. 행동 학습의 구조
Fig. 3. Structure of behavior learning

3.3 CSVM(Cascade Support Vector Machine)

Cascade Support Vector Machine(CSVM)은 많은 양의 독립적인 문제들을 분할해 계산한다는 부분에 기반을 둔 학습 방법이다. 독립적인 문제에 대한 결과 값들이 서로 결합되어서 hierarchical한 방법으로 다음 층으로 이동하게 되어 있는 다음과 같은 방법은 각각의 문제가 서로 독립이라고 가정하고 출발한다[8][9].

Training Data(TD)에 대해서 각각의 Support Vector(SV)는 독립적으로 계산을 수행한다. 2개의 Support Vector가 결합된 그 결과 값 들은 다음 층의 Support Vector의 입력 값으로 들어가게 되며, 각 층에서 가장 적합한 Support Vector를 계산한다. 최종적으로 도달한 최하위 층의 Support Vector는 global convergence를 계산하여 유효한 Support Vector를 생성하였을 경우에는 그 값이 최적의 Support Vector가 되어 결과 값을 얻을 수 있으며, 그렇지 못할 경우에는 다시 첫 번째 층의 Support Vector와 비교하는 과정을 반복하여 수행하여 최적의 Support Vector를 구해내게 된다(그림 4).

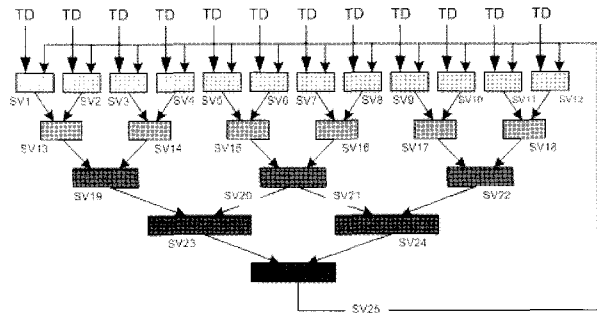


그림 4. 종속적인 SVM의 구조
Fig. 4. Structure of Cascade SVM

3.4 제안한 Q-learning + Cascade SVM 알고리즘

본 시스템에서 로봇의 행동에 대한 보상이나 벌칙은 주로 이전 행동의 영향에 의하여 받은 것이다. 또한 현재 받은 보상이나 벌칙은 바로 다음에 이어지는 상태에는 별로 영향을 미치지 않는다. 따라서 학습은 보상이나 벌칙을 받은 시점에서 과거의 행동에 대하여 수행하는 지연보상이 있는 Q-학습법을 사용하였다. 다음의 표는 본 논문에서 제안한 Q-learning + Cascade SVM 알고리즘이다.

- Step 1. 모든 상태 s_1 과 행동 a 에 대하여 $Q(s_1, a)$ 을 임의의 값(일반적으로 0)으로 초기화한다.
- Step 2. 현재의 상태 s_1 을 인식한다.
- Step 3. (상태, 행동) 규칙에 따라 행동 a 를 선택한다.
- Step 4. 주어진 환경에서 행동 a 를 수행하고, 다음상태를 s_1' 즉각적인 보상을 r 로 놓는다.
- Step 5. 다음 단계에서 즉각적인 보상 대신에 현재 상태 s_1 , 다음 상태 s_1' 과 또 다른 로봇의 현재 상태 s_2 , 다음 상태 s_2' 를 데이터베이스화 한다.
- Step 6. 모든 로봇의 상태 s_n 을 분류하여 올바른 행동, 올바르게 못한 행동으로 분류를 한다.

$$f(S) = \text{sgn}(\langle S, X \rangle + b) \quad (2)$$

- Step 7. 올바른 행동 상태를 만들기 위하여 현재 상태에 대한 데이터베이스를 SVM 분류기를 이용하여 최적화시킨다.

$$\text{maximize } \min \{ \| S - S_i \| | \langle S, X \rangle + b, i = 1, \dots, n \} \quad (3)$$

$b \in R$

- Step 8. SVM 분류기를 생성해 $TD_1, TD_2, \dots, TD_{12}$ 을 학습시킨다. 첫 번째 층에서 생성된 support vectors의 집합을 $SV_1, SV_2, \dots, SV_{12}$ 이라고 한다.
- Step 9. $SV_1, SV_2, \dots, SV_{12}$ 을 결합시키고 결합된 데이터 집합을 학습시켜 $SV_{13}, SV_{14}, \dots, SV_{18}$ 을 생성한다. 3번째 층과 4번째 층에서도 동일한 과정을 거쳐서 $SV_N (N = 1, 2, \dots, 25)$ 을 생성한다.
- Step 10. $TD_2, TD_3, TD_4, \dots, TD_{12}$ 에 SV_1 을 추가하고 $TD_1, TD_3, TD_4, \dots, TD_{12}$ 에 SV_2 을 추가한 후 step2와 동일한 방법으로 학습시킨다. 여기서 생성된 support vectors의 집합을 $SV'_1, SV'_2, SV'_3, \dots, SV'_{25}$ 이라고 한다. 만약 SV_1 과 SV'_1 , SV_2 과 SV'_2 , SV_3

와 SV'_3, \dots , 그리고 SV'_{25} 와 SV'_{25} 이 차이점이 없거나 차이점이 있는 집합의 각 요소들이 고정된 값이 나오거나, 혹은 그 차이가 주어진 값 ϵ 보다 작을 때는 최종적으로 나오는 결과 값은 SV'_{25} 가 되고 그때 SVM 함수를 얻을 수가 있다.

Step 11. 만약 SV_1 와 SV'_1, SV_2 와 SV'_2, SV_3 와 SV'_3, \dots , 그리고 SV'_{25} 와 $9SV'_{25}$ 이 위의 조건에 해당되지 않으면 Step 8로 돌아간다.

Step 12. 최적화 시킨 상태의 행동들을 대입하여 s_1, a, s'_1, r 로부터 행동 규칙을 갱신해 나간다.

$$Q_{t+1}(s_1, a) = (1 - \alpha_t)Q_t(s_1, a) + \alpha_t(r + \max_{a' \in A} Q_t(s'_1, a')) \quad (4)$$

식에서 α_t 는 학습률이고, Y 는 0과 1사이의 고정된 감쇠 계수이다.

여기서 S_N 이 가질 수 있는 값을 12가지라 하고, 12가지의 값과 a_N 이 가질 수 있는 값도 최대 12가지라고 하면 각각의 값이 Training Data (TD)가 되어서 입력 값으로 들어가고 25개의 SVM 분류기로부터 얻어진 25개의 Support Vector를 통해서 최종적으로 가장 적절한 값을 결정하게 된다.

4. 분산유전알고리즘에 의한 행동 진화

4.1 분산유전알고리즘(Distributed genetic algorithm)

분산유전알고리즘은 유전 알고리즘의 변형된 것으로서 다음의 세 가지가 있는데 첫째는 개체군을 여러 개의 군으로 나누어 다른 컴퓨터에서 진화하면서 각각의 개체군을 통합하는 방법이고, 둘째는 하나의 개체(agent)가 하나의 염색체가 되며 각 개체 간 통신에 의한 방법 등으로 일괄적이 아닌 분산적으로 진화하는 방법, 셋째는 염색체를 여러 개의 부분으로 나누어 개체(agent)에 할당하고 임무 수행 후 다시 개체를 합쳐 하나의 염색체로 재구성하는 방법이다.

본 논문에서는 하나의 개체(agent)가 하나의 염색체가 되며 각 개체 간 통신에 의한 방법 등으로 분산적으로 진화하는 방법을 사용하였다. 이 방법은 진화의 대상인 염색체가 하나의 로봇이 됨으로서 여러 대의 로봇으로 구성되어 있는 자율분산로봇시스템에 실제적으로 적용하여 각각의 개체인 로봇이 시스템의 목적에 맞도록 진화를 시킬 수 있는 장점이 있다.

단순 유전 알고리즘에서는 적합도의 평가, 선택, 교배 및 돌연변이의 과정이 일괄적으로 이루어진다. 그러나 분산유전알고리즘은 이러한 연산이 각 개체에 대하여 분산적으로 이루어진다. 이것은 진화 알고리즘의 연속세대 모델에 해당하는 것으로 각각의 개체(본 논문에서는 로봇이 됨)는 적합도의 평가 능력과 선택, 염색체의 교배 및 돌연변이의 기능을 갖추어야 한다. 이러한 진화의 과정에 의해 로봇은 다른 환경에서 획득된 우수한 로봇의 염색체를 받아들여 자신의 수행능력을 향상시킨다. 이것은 다수의 로봇에 의한 협조 학습(진화)의 과정으로 볼 수 있다.

4.2 염색체(Chromosome)

로봇이 진화의 대상으로 하는 것은 자신이 가지고 있는

염색체이다. 본 논문에서는 로봇이 현재까지 학습한 데이터인 Q-테이블의 값을 염색체로 하였다. 이 Q-테이블의 값은 로봇이 환경에 대응하여 학습한 결과로서 로봇마다 각자 학습한 다른 값을 가지고 있으며 진화의 대상으로 하기에 적당하다.

4.3 선택(Selection)

환경에 대하여 평가도 받아보지 못한 로봇이 선택되는 것을 방지하기 위하여 교배 후 최소한 일정한 주행 시간이 지난 로봇에 대하여 선택 될 수 있는 자격을 부여한다. 만약 어떤 로봇이 자신보다 우수한 로봇을 만나면 그 로봇을 선택하여 유전자를 받아오고 자신의 유전자와 교배를 하여 새로운 유전자를 만들어낸다.

4.4 교배(Crossover)

염색체가 Q-테이블이기 때문에 하나의 상태와 그 상태에서 취할 수 있는 행동의 집합을 하나의 유전자로 하였다. 따라서 유전자의 총 수는 로봇이 가질 수 있는 총 상태의 수가 된다. 일반적으로 두 개의 부모개체를 선택하여 교배하면 새로 생기는 자식의 개체는 두 부모의 특성을 함께 가지게 된다. 이때 두 부모개체의 형질은 새로 생겨난 자식들에게 유전됨으로서 일단 선택이 된 부모의 염색체는 소실되지 않고 두 개의 자식에게 나누어 유전된다. 그러나 본 논문에서 사용하는 분산유전알고리즘은 하나의 로봇은 선택에 의해 가져온 염색체와 자신의 염색체를 합쳐 새로운 하나의 염색체를 재생산하여 자신의 염색체로 치환하므로 두 부모의 염색체 중 절반은 소실된다. 따라서 이러한 교배에 의하여 우수한 개체가 소실될 가능성도 존재한다. 이러한 점을 보완하기 위하여 본 논문에서는 개선된 교배방법을 제안한다.

제안된 교배의 방법은 기본적으로 일정교배와 유사하다. 그러나 강화학습의 특성을 살리기 위하여 임의로 발생한 0과 1의 마스크를 사용하여 교배를 하는 기존의 방법과는 달리 각 유전자는 현재까지 학습한 횟수를 저장하고 있어서 이 횟수에 비례하여 두 로봇의 유전자 중 하나의 유전자를 택한다. 결국 학습이 많이 된 유전자에 대하여 선택될 확률을 높여서 좋은 유전자의 소실을 막을 수 있도록 하였다.

부모 개체의 두 염색체를 g (유전자)와 q (학습된 횟수)의 상으로 표현하면 로봇 1과 로봇 2의 염색체는 (5)식과 같이 나타낼 수 있다.

$$\begin{aligned} (\vec{g}^1, \vec{q}^1) &= ((g_1^1, \dots, g_n^1), (q_1^1, \dots, q_n^1)) \\ (\vec{g}^2, \vec{q}^2) &= ((g_1^2, \dots, g_n^2), (q_2^2, \dots, q_n^2)) \end{aligned} \quad (5)$$

단 n 은 총 유전자의 개수이다.

이때 새로운 교배 방법에 의하여 생성되는 염색체는 (6)식과 같이 나타낼 수 있다.

$$(\vec{g}, \vec{q}) = ((g_1^{s_1}, \dots, g_n^{s_n}), (q_1^{s_1}, \dots, q_n^{s_n})) \quad (6)$$

단 $s_i = \begin{cases} 1 & p_i < \frac{q_i^1}{q_i^1 + q_i^2}, i=1, \dots, n \text{ 이고 } p_i \text{는 } 0 \text{과} \\ 2 & \text{else} \end{cases}$

1사이의 임의의 난수이다. 즉 유전자의 학습된 회수를 고려하여 부모 1과 2의 염색체를 유전 받는다.

4.5 적합도 함수(Fitness function)

적합도 함수는 진화의 방향을 결정하는 가장 중요한 parameter이다. 실제적으로 이 적합도 함수에 의하여 로봇들이 원하는 행동이나 협조행동을 하도록 진화해 간다. 뿐만 아니라, 적합도의 값은 서로 다른 로봇을 선택하는 기준이 된다. 본 논문에서는 협조 탐색의 문제로서 충돌을 피하면서 많은 물체를 획득하는 것을 목표로 하고 있으므로, 물체를 획득하였을 경우 적합도가 상승하고, 장애물이나 로봇에 충돌하였을 경우 적합도가 떨어진다. 여기에서 로봇의 적합도는 최종 주행시간 동안 받은 보상이나 벌칙에 의해 다음 (7)식과 같이 표현할 수 있다.

$$fitness = \alpha x + \beta y + \gamma z \quad (7)$$

여기서, x는 보상의 수이고, y는 벌칙의 수, z는 로봇이 취하는 행동에 대하여 소비되는 에너지의 양을 각각 나타낸다. 그리고 α, β, γ 는 중요도를 나타내는 비례상수이다.

적합도 함수에 의하여 로봇의 진화의 추이를 변화시킬 수 있다. 예를 들어 (7)식에서 α 의 값을 상대적으로 크게 하면 충돌이나 에너지 소비를 작게 하는 개체보다는 물체 획득에 뛰어난 개체가 많이 생겨난다. 그러나 초기에 학습의 영향이 클 때에는 진화의 방향도 세 가지 항목에 대하여 비슷하게 이루어지지만 시간이 지남에 따라 학습의 영향이 줄어들게 되므로(온도계수 T의 감소에 의해) 점점 진화의 영향이 증대된다.

5. 시뮬레이션 결과

본 논문에서는 자율이동로봇 시스템의 행동학습과 진화를 위하여 군집 로봇의 협조탐색 문제의 하나인 다수 로봇에 의한 물체획득 문제로 설정하였으며 실험을 위하여 다음과 같은 환경을 가정하였다. 25대의 로봇과 50개의 장애물, 그리고 50개의 목표물이 있으며, 작업공간은 100×100(로봇의 크기는 2×2, 목표물의 크기 2×2), 통신반경은 25×25, 로봇이 1sec 동안 이동할 수 있는 거리는 2×2 크기로 가정하였고, 탐색시간은 1회당 총 60sec로 총 100번에 걸쳐서 진행하였다. 또한 작업 대상인 물체 및 장애물은 작업 공간 내에 골고루 퍼져 있고, 모든 로봇은 작업공간에서 다른 로봇과의 거리를 충분히 유지하도록 흩어진 후 작업을 수행한다고 가정한다. 제안된 방법들의 유효성의 검증을 위하여 시뮬레이션에서는 학습과 진화를 하지 않을 경우, 학습만을 수행할 경우, 진화와 학습을 동시에 수행할 경우에 대하여 수행하여 결과를 비교하였으며, 적합도 함수는 α 의 비중을 크게 한 경우와 β 의 비중을 크게 한 경우에 대하여 각각 진화의 추이를 비교하였다.

그림 5는 시행회수에 따른 적합도의 총합을 나타낸 그림이다. 그림에서 ①은 학습과 진화를 하지 않을 경우, ②는 학습만을 수행할 경우, ③은 진화와 학습을 동시에 수행할 경우를 각각 나타낸다. 여기서 α, β 는 10이라고 가정한다. 그림 6은 시행회수에 따른 물체의 획득 수를 나타낸 그림이고, 그림 7은 시행회수에 따른 장애물과의 충돌 회수를 나타낸 그림이다. 이들 그림에서 ①과 ②선은 (7)식의 적합도 함수에서 진화에 영향을 미치는 비례상수의 크기가 ①의 경우 $\alpha=20, \beta=1$ 인 경우이고, ②의 경우는 $\alpha=1, \beta=20$ 인 경우이다. 결과에 의하면 획득한 물체에 대하여 적합도 함수에서 α 의 비중을 크게 하면 장애물 회피보다는 물체의 획득에 더 우수한 능력을 나타냈으며, 반면 β 의 비중을 크

게 하면 그 반대의 경우가 나타났다. 이러한 차이는 적합도가 다른 로봇의 선택의 기준이 되기 때문에 적합도의 특성에 맞는 로봇이 주로 선택된 결과이다. 이때 뚜렷한 진화의 영향은 학습이 어느 정도 진행된 후에 나타난다.

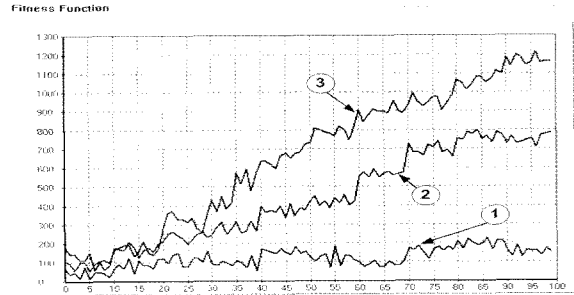


그림 5. 시행회수에 따른 적합도 함수의 변화($\alpha, \beta=10$)
Fig. 5. Relationship between total fitness variation and iteration numbers($\alpha, \beta=10$)

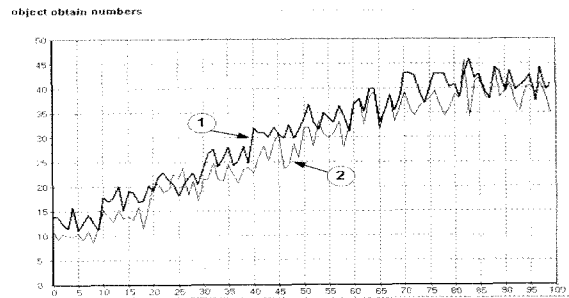


그림 6. 시행회수에 따라 획득한 물체의 수
Fig. 6. Relationship between object obtain numbers and iteration numbers
(① : $\alpha=20, \beta=1$ ② : $\alpha=1, \beta=20$)

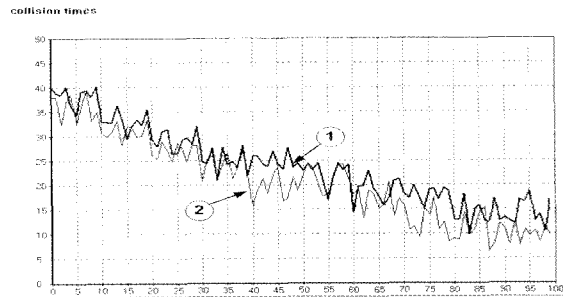


그림 7. 시행회수에 따른 충돌회수
Fig. 7. Relationship between collision times and iteration numbers
(① : $\alpha=20, \beta=1$ ② : $\alpha=1, \beta=20$)

6. 결 론

본 논문에서는 다수의 로봇으로 구성된 군집 로봇시스템에서 로봇의 행동학습 및 진화를 위하여 Cascade SVM을 기초로 한 강화학습과 분산유전알고리즘을 도입한 방법을 제안하였다. 각각의 로봇은 센싱을 통하여 주변을 인식하여

자신의 행동을 결정하며, 이때 Cascade SVM의 분류기를 통해 보상을 하는 Q-학습법을 제안하며 적용하였고, 지역적 통신시스템을 이용하여 시스템의 목적에 맞도록 진화해 나가는 방법을 사용하였다. 제안된 방법은 강화학습의 즉각적인 보상보다 더 정밀하게 옳고 그른 행동을 구분 지음으로써 로봇의 행동을 정확하게 학습시킬 수가 있었다. 본 논문에서는 로봇에게 완전한 프로그램을 만들어 주는 대신 동적으로 변화하는 환경에 대하여 유연하게 대처할 수 있는 행동이 발현되고 진화해 나갈 수 있는 시스템을 구현하여 협조행동을 실현하였다.

향후 과제로는 군집 로봇 시스템 상에서 적용하기 위해 로봇의 수를 수백 대 혹은 수천 대 로 증가시켜서 실험해 봄으로써 본 알고리즘의 유효성을 다시 한 번 판단 해 보아야 할 것이다. 또한 실제로 동적인 환경에서 본 알고리즘을 직접 모델링 해 나가는 과정도 필요 할 것이다.

[주] 본 논문은 2008년도 한국지능시스템학회 추계학술대회에서 발표된 우수논문[1]으로 개정된 것임.

참 고 문 헌

[1] 서상욱, 양현창, 심귀보, "Q-learning과 Cascade SVM을 이용한 군집로봇의 행동학습 및 진화," 2008년도 한국지능시스템학회 추계학술대회 학술발표논문집, 18권, 제2호, pp. 257~260, 2008. 10. 31~11. 1.

[2] 심귀보, 이동욱, "군집 로봇의 군 행동을 위한 통신 모델과 이론적인 해석," 퍼지 및 지능시스템학회 논문지, vol 16, no 1, pp. 8-17, 2006.

[3] M.J. Mataric, "Behavior-based control: Examples from navigation, learning, and group behavior," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 9, no. 2, pp. 323-336, 1997.

[4] 이동욱, 심귀보, "인공면역계 기반의 자율이동로봇 군의 협조행동전략 결정," *대한전자공학회 논문지*, 제35권 S편 제3호, pp. 102-109, 1998. 3.

[5] Adam T. Hayes, Alcherio Martinoli and Rodney M. Goodman, "Swarm robotic odor localization: Off-line optimization and validation with real robots," *Robotica*, vol. 21, pp. 427-441, 2003.

[6] 이호근, 김명훈, 이지근, 정성태, "SVM-SMO와 Pan-Tilt 웹카메라를 이용한 실시간 얼굴 추적과 얼굴 인식," *한국정보과학회*, vol. 31, no 2, pp. 679-681, 2004.

[7] H.B. Jun, K.B. Sim, "Emergence of Cooperative Behavior based on Learning and Evolution in Collective Autonomous Mobile Robots," *Journal of Electrical Engineering and Information Science*, vol. 3, 1998.

[8] Hans Peter Graf, Eric Cosatto, Leon Bottou, Igor Durdanovid, Vladimir Vapnik, "Parallel Support Vector Machines:The Cascade SVM," *Springer*, 2004.

[9] Jing Yang, "An Improved Cascade SVM Training Algorithm with Crossed Feedbacks," *Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences*, 2006.

저 자 소 개



서상욱(Sang-Wook Seo)
 2007년 : 중앙대학교 전자전기공학부 공학사
 2009년 : 중앙대학교 대학원 전자전기공학부 공학석사
 2009년~현재: 중앙대학교 대학원 전자전기공학부 박사과정

관심분야 : Machine learning, Group behavior and control of swarm robot, Multi agent robotic system 등



양현창(Hyun-Chang Yang)
 2006년~현재 : 중앙대학교 대학원 전자전기공학부 박사과정

관심분야 : 군집로봇 시스템, 유비쿼터스 센서 네트워크, 스마트 홈 등



심귀보(Kwee-Bo Sim)
 1990년 : The University of Tokyo 전자공학과 공학박사
 1991년 ~ 현재 : 중앙대학교 전자전기공학부 교수

[제19권 제1호 (2009년 2월호) 참조]

E-mail : kbsim@cau.ac.kr
 Homepage URL : http://alife.cau.ac.kr