

자율이동로봇의 동적 편대 형성과 장애물 회피를 위한 신경망 구조 및 강화학습

민석기, 신석영, 강 훈
중앙대학교 공과대학 전자전기공학부

A Neural Network Model and Reinforcement Learning for Dynamic Formation Moving and Obstacle Avoidance of Autonomous Mobile Robot

Suk-Ki Min, Suk-Young Shin and Hoon Kang
School of Electrical & Electronic Engineering
College of Engineering, Chung-Ang University

Abstract - The objective of this paper is, based upon the principles of artificial life, to induce emergent behaviors of multiple autonomous mobile robots which form from simple local rules to complex global intelligence. Here, we propose an architecture of neural network learning with reinforcement signals which perceives the neighborhood information and decides the direction and the velocity of movement as mobile robots navigates in a group. As results of the simulations, the optimum weights are obtained in real time, which not only prevent from the collisions between agents and obstacles in the dynamic environment, but also have the mobile robots move and keep in various patterns.

1. 서 론

최근의 복수의 자율 이동 로봇 시스템에 대한 관심은 개체간 국부적 상호 작용에 집중되어 있다. 인공생명[1]은 지능을 하나의 독립적인 것으로 보는 기존의 인공지능에 비해 군 지능이 독립적인 것이 아닌 개체간의 상호작용으로부터 이루어진다는 것을 강조하고 있으며 이러한 상호작용으로부터 전역적 행동을 이해할 수 있다고 주장한다. 인공생명 기법을 이용한 많은 연구 결과들이 발표되고 있다.[2,3,4,5] 인공생명에 의한 로봇은 문제를 수행함에 있어서 사전에 명백히 프로그래밍 되어있지 않더라도 예측하지 못한 문제가 발생했을 경우 이에 적응 및 학습을 통해 대처해 나가는 능력이 있다. 이를 구현하기 위해 자연 생물계의 메커니즘을 모방한 유전자 알고리즘, 유전자 프로그래밍, 진화 전략, 진화 프로그래밍과 같은 진화연산과 신경 회로망, 강화학습, 면역 시스템과 같은 적응행동이 많이 연구되고 있다. 본 논문에서는 자율 이동 로봇의 군 행동을 제어하는 문제를 다룬다. 복수 이동 로봇 에이전트를 이용하여 이들이 다른 개체와 충돌하지 않으면서

무리를 지어 이동하는 복잡한 행동이 단순한 지역적 상호작용에 의해 구현 될 수 있음을 보여주고 인공생명기법을 이용해 원시적인 상태에서 효율적인 전략적 규칙을 갖는 상태로 진화해가는 것을 보여주고자 한다. 여기서 제안한 군 행동에 대한 제어구조의 유효성은 컴퓨터 시뮬레이션을 통해 테스트하였다. 관련된 연구는 군 지능에 관한 창발적인 행동을 실험한 결과들이 있다.[6,7]

II. 본 론

2.1 Action Network의 구성

이동 로봇은 센서로부터 입력을 받으며 입력 변수는 다른 로봇이나 장애물의 위치 벡터이다. 센서는 센싱 반경을 가지고 있으며 그 영역은 무리 짓기 영역과 충돌회피영역으로 나누어져 있다. 자율 이동 로봇의 출력변수는 개체가 갈 수 있는 거리(속력)과 회전하는 방향(방향벡터), 즉 속도가 된다.

2.1.1 속력의 결정

속력 $|v_i(t)|$ 의 결정은 무리를 이루어 다른 로봇과 적절한 거리를 유지하며 진행하도록 정한다. $\Delta s(t)$ 를 다른 로봇과 적절한 거리를 유지하기 위해 자신이 가감해야할 거리라고 하면 $\Delta s(t)$ 에 의한 가속도로부터 새로운 속력을 얻을 수 있다:

$$|v_i(t+1)| = |v_i(t) + a_i(t)| \quad (1)$$

각 자율 이동 로봇은 (1)식에 의해 새로운 속력을 구한다. 여기서, 고려된 사항은 각 자율 이동 로봇은 평균 속력 $|v^*|$ 을 유지하고 최대 속력과 최소 속력에 의해 제한된 결과를 갖도록 하는 것이다.

2.1.1 방향벡터의 결정

방향 벡터를 결정하는 전략은 센싱 영역 안에 어떤 개체가 있는냐에 따라 3가지가 있다. 임의

의 벡터방향 P_i 가 방향 벡터 d_i 의 오른쪽으로 향하면 우벡터 P_i^+ , 왼쪽을 향하면 좌벡터 P_i^- 로 정의한다. 만약 센싱 범위 안에 다른 개체가 존재하지 않는다면 로봇은 한 스텝 안에서 움직일 수 있는 최대 제한각도 안에서 자유롭게 움직인다.

1.1 무리짓기 운동

센싱 영역안에 다른 로봇이 들어왔을 경우 무리를 만들기위해서 진행하는 방향을 결정하는 것은 다른 로봇과의 상대적 거리와 방향의 관계임을 알 수 있다. 식(2)에서 무리짓기운동에서 방향벡터를 결정하며 주연산자로서 벡터 합을 사용함으로써 시스템의 출력은 다른 로봇들이 밀집한 방향으로 되어 무리를 형성할 확률이 높아진다.

$$d_i(t+1) = d_i(t) + \sum_{j=1}^{N-1} (w_j^q(t) \cdot P_j^q(t) + O_j^q(t))$$

$$O_j^q(t) = \begin{cases} O_j^q(t) & \text{if } P_j^q(t) = [P_j^q(t)]_i^- \\ -O_j^q(t) & \text{if } P_j^q(t) = [P_j^q(t)]_i^+ \end{cases}$$

$$P_j^q(t) \cdot O_j^q(t) = 0 \quad (|O_j^q(t)| = 1) \quad (2)$$

여기에서 $w_j^q(t)$ 는 로봇에 대한 하중값, $P_j^q(t)$ 는 시간 t 에서 j 번째 로봇이 센싱 반경 안에 들어온 j 번째의 다른 로봇을 가리키는 위치 벡터, $O_j^q(t)$ 는 이에 수직인 단위 벡터이고 N 은 센싱 반경 안에 들어온 로봇의 개수이다. 그림 4는 i 번째 로봇의 방향 벡터 $d_i(t)$ 를 결정하는 시스템의 구성도를 나타낸다. 하중값 $w_j^q(t)$ 가 방향 벡터를 결정하는 변수이므로 적절한 값을 선택하는 것이 매우 중요하다.

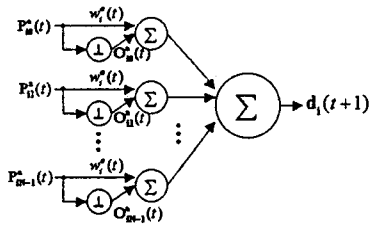


그림 1. i 번째 로봇의 시스템의 구성도

1.2 장애물(경계면) 회피 운동

현재 이동 로봇이 무리짓기 운동을 하고있지 않고 센싱영역내에 장애물(경계면)이 들어왔을 경우 하는 운동이며 시스템의 입력으로 자신의 상태, 회전 비트 OB (orientation bit)가 들어간다. 회전 비트(OB)는 현재 자신이 회전하는 방향을 나타내며 오른쪽으로 회전($OB=1$), 왼쪽으로 회전($OB=-1$), 장애물 없음 또는 우선권이 높은 장애물의 변경($OB=0$)의 3가지 상태가 있다. 임의 로봇이 장애물을 만났을 때, 진행 방향을 결정하는 것은 다음의 식으로 구한다:

$$d_i(t+1) = d_i(t) + \min_{j=0}^{M-1} (w_j^q(t) \cdot P_j^q(t) + O_j^q(t))$$

$$O_j^q(t) = \begin{cases} O_j^q(t) & \text{if } OB=1 \\ -O_j^q(t) & \text{if } OB=-1 \\ O_j^q(t) & \text{if } OB=0 \text{ \& } P_j^q(t) = [P_j^q(t)]_i^- \\ -O_j^q(t) & \text{if } OB=0 \text{ \& } P_j^q(t) = [P_j^q(t)]_i^+ \end{cases}$$

$$P_j^q(t) \cdot O_j^q(t) = 0 \quad (|O_j^q(t)| = 1) \quad (3)$$

여기에서 $w_j^q(t)$ 는 장애물(경계면)에 대한 하중값, $P_j^q(t)$ 는 시간 t 에서 j 번째 로봇이 센싱 반경 안에 들어온 j 번째 장애물을 가리키는 위치 벡터, $O_j^q(t)$ 는 이에 수직인 단위 벡터이며 M 은 센싱 반경 안에 있는 장애물의 개수이다. 로봇은 가장 가까운(우선권이 높은) 장애물(경계면)에 대해 하중값과 위치 벡터에 따라 진행 방향을 결정한다. 회전 비트는 수직 단위 벡터를 결정하는데 사용되며 그 영향으로 로봇은 보다 안전하고 효율적으로 장애물을 피한다. 회전 비트는 단지 장애물(경계면) 회피 운동에서만 유효하다.

1.3 충돌 회피 운동

충돌 회피 영역 안에 다른 로봇이나 장애물이 들어왔을 경우이며 가장 우선권이 높은 행동이다. 따라서, 다음과 같이 위험 비율을 정의해서 가장 안전한 방향을 선택하였다.

$$\text{danger ratio} = 1 - |P_{ij}|/AR \quad (4)$$

(5)식에서 AR 은 그림 1에서 보여준 충돌 회피 영역의 반경이며 P_{ij} 는 다른 로봇이나 장애물(경계면)을 가리키는 위치 벡터이다. 그림 2는 위험 비율의 적용 예를 보여주며 충돌 회피 영역을 임의의 크기(센서의 개수, n)로 나눈 부분을 세포(cell)라고 부른다. 각 세포는 하나의 진행 방향을 대표하고 있으며 (5)식에서 구한 위험 비율 값을 가지고 있다. 가장 작은 값을 가진 세포가 다른 로봇이나 장애물(경계면)과 충돌할 확률이 가장 적은 방향이며 다음 스텝에서의 진행 방향이 된다. 로봇은 한 스텝에서 회전할 수 있는 최대 제한각을 고려하여 세포를 선택한다.

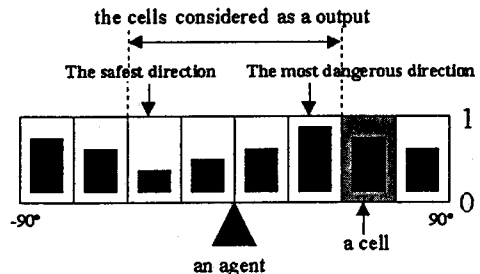


그림 2. 위험 비율의 예

2.2 자율 이동 로봇의 학습에 의한 진화

로봇 개체의 주변 환경은 동적으로 변화하므로 이에 따라 자신의 행동을 조정해서 원하는 목적을 이루는 것이 효율적이고 지능적이라고 말할 수

있다. 따라서 인공생명체에 의한 로봇은 문제를 수행함에 있어서 사전에 명백히 프로그래밍 되어있지 않더라도 예측하지 못한 문제가 발생했을 경우 이에 적응 및 학습을 통해 대처해 나가는 능력이 있어야 한다. 적응 행동과 진화 연산을 구현하는 기법으로 생물체의 적응 메커니즘을 인공적으로 적용한 구조가 관심을 끌고 있다. 신경 회로망[8,9]은 동물의 신경 세포를 기반으로 한 구조이며 강인한 학습 능력을 가지고 있고, 강화 학습은 행동의 학습을 통해 상태와 행동을 연결시켜주는 학습 구조로 Sutton의 Adaptive Heuristic Critic 구조[11]와 Watkins의 Q-학습법[12] 등이 있다. 본 연구에서는 교사 학습에 의한 진화의 예로 자율 이동 로봇군이 특정한 편대의 모양을 유지하며 무리를 지어 이동하는 것을 목표로 설정하였으며 학습을 통해 하중값을 조정함으로써 원하는 목적을 달성하도록 하였다. 편대의 모양은 그림 3과 같으며 자신의 선두는 특별히 정해져 있지 않고 임의의 로봇이 될 수 있으며 언제든지 변할 수 있다. 선두는 가장 근거리의 로봇으로 선택한다.

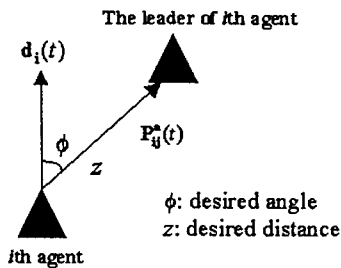


그림 3. 원하는 무리의 부분적 형태

- (1) 거리 z 의 보정: 속력을 조정하여 원하는 거리를 유지하도록 한다.

$$\Delta s(t) = \begin{cases} 0 & \text{if } |P_g^*(t)| \leq z \pm a \\ |P_g^*(t)| - z & \text{if } |P_g^*(t)| > z \pm a \end{cases} \quad (5)$$

(a 는상수)

- (2) 각도 ϕ 의 보정: 선두와 임의의 각을 이루도록 하중값 w_i^* 를 조정한다. 하중값을 수정하는 규칙은 제곱 오차 최소화 방법에 따라 다음과 같다:

$$\varepsilon_i^2(t) = \frac{1}{2}(\phi - \theta_g(t))^2 \quad (6-1)$$

$$w_i^*(t+1) = w_i^*(t) - \alpha \cdot \varepsilon_i(t) \quad (6-2)$$

- (3) 강화 신호에 의한 전략: 상태에 따른 보상과 벌칙 신호를 주어 로봇이 보다 효율적인 방향으로 원하는 목적을 이루도록 하였다. 강화 신호 $r_i(t)$ 평가함수는 다음과 같다:

$$r_i(t) = \begin{cases} 1 & \text{if } \Delta s(t) < 0 \\ 0 & \text{if } \Delta s(t) = 0 \\ -1 & \text{if } \Delta s(t) > 0 \end{cases} \quad (7)$$

따라서, 하중값의 수정 규칙은 (6)식과 (7)식에서 다음과 같이 정리된다.

$$w_i^*(t+1) = w_i^*(t) - \alpha(r_i(t) + \lambda \varepsilon_i(t)) \quad (8)$$

여기서, α 는 학습률, λ 는 비례상수이다. 그림 4

는 신경 회로망의 학습 구조를 사용한 자율 이동 로봇 개체의 구성도이다.

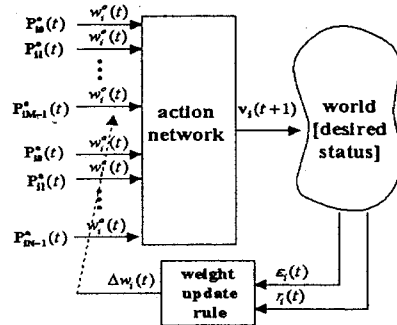


그림 4. 자율 이동 로봇의 행동-학습 구조

2.3 모의 실험에 의한 검증

2.3.1 모의 실험 결과

아래의 모의실험에서 센싱 반경 100, 충돌 회피 반경 20, 평균 속도 2를 공통적으로 설정했다. 그림 5는 학습을 사용하지 않고 고정된 하중값을 사용했을 경우 무리지어 이동하는 것을 보였다. 이 경우 로봇의 수 10개, 하중값 0.04,로 정하였다. 그림 6은 로봇이 장애물과 경계면을 피해 움직이는 궤적을 보였다. 이 경우 장애물에 대한 하중값 0.03으로 정하였다. 그림 7은 8개의 자율 이동 로봇군이 장애물과 경계면을 피하면서 학습을 통해 특정한 형태로 무리지어 이동하는 과정을 보여주고 있다. 원하는 거리 $z=40$, 원하는 각도 $\phi=30^\circ$, 최고 속도 4, 최저속력 1, $\alpha=0.001$, $\lambda=1$, $a=6$ 으로 정하였다. 그리고 장애물의 하중값은 0.02로 하였다. 그림 8은 원하는 거리 z 를 40으로 고정시키고 원하는 각도 ϕ 를 변화시켰을 때 나타나는 편대의 모습이다.

2.3.2 성능 지수

우리는 원하는 무리를 유지하게 해주는 최적의 하중값 w_i^* 이 존재함을 추론할 수 추론할 수 있다. 이 값은 원하는 위치에서 최적의 값이며 다른 위치에서는 최적의 값이 아니다. 이 최적의 하중값은 다음과 같다:

$$w_i^* = \frac{1}{z \cdot \tan \phi} \quad (9)$$

따라서, 원하는 편대 형태의 무리가 형성된다면 선두를 제외한 로봇들은 선두가 무작위 움직임을 하므로 최적의 하중값 w_i^* 근처에서 진동할 것이다. 따라서 하중값 궤적 곡선을 조사하여 시스템의 성능 지수를 확인할 수 있다. 그림 8은 충돌 회피 영역을 20으로 설정하였을 때 원하는 거리 z 및 원하는 각도 ϕ 대 최적의 하중값 w_i^* 에 대한 그래프를 보여주고 있다. z 와 ϕ 가 작은 경우, 즉 가까운 거리를 유지하면서 선두를 일직선으로 뒤따라가기 위해서는 하중값 w_i^* 이 급격히 증가해야 한다는 것을 알 수 있다.

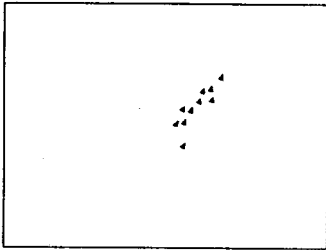


그림 5. 고정된 하중값에 대한 무리짓기

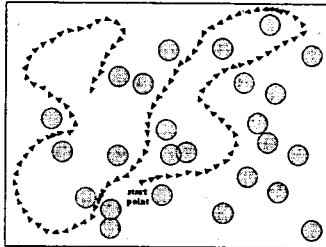


그림 6. 로봇이 장애물을 피하는 궤적

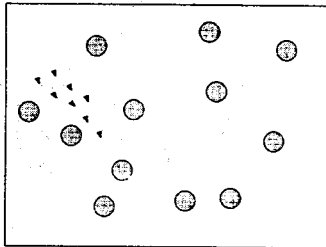


그림 7. 특정한 형태를 유지하며 무리짓기

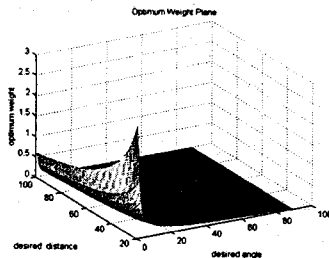


그림 8. 최적의 하중값 곡선

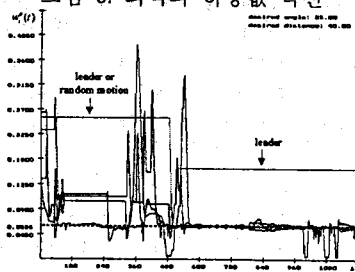


그림 9. 자율 이동 로봇의 하중값 궤적

그림 9는 6개의 로봇이 원하는 거리를 40, 원하는 각도를 25°로 설정했을 때 시간에 따른 각 로봇의 하중값 w_i 궤적을 보여주고 있다. 시간 550이후 완전한 하나의 무리를 만들었음을 알 수 있다.

III. 결 론

자율 이동 로봇의 군행동을 통해 개체간의 지역적 상호작용으로부터 전역적 군행동이 유발되는 것을 살펴보았다. 제안된 자율 이동 로봇의 구조는 신경 회로망의 학습과 강화 신호를 사용하여 인공 생명의 효율적인 전략을 가지고 환경에 따라 적응 행동을 이루는 것을 보여준다. 신경망 기반의 분산된 자율 이동 로봇의 행동은 동적 환경에서 움직이는 다른 개체에 대해서도 적응하여 학습을 통한 최적의 하중값을 스스로 선택해나가는 것을 확인하였다.

[참고문헌]

- [1] C. Langton, "Artificial Life" in *Artificial Life*, C. Langton (ed.), Addison-Wesley, pp.1-47, 1989.
- [2] R. A. Brooks, "Behavior Humanoid Robotics", *Proc. of Int. Conf. on IROS*, pp.1-8, 1996
- [3] M. J. Mataric, "Designing Emergent Behaviors: From Local Interactions to Collective Intelligence", *Proc. 2nd Int. Conf. on Simulation of Adaptive Behavior*, pp.432-441, 1993.
- [4] C. R. Kube and H. Zhang, "Collective Robotic Intelligence", *Proc. 2nd Int. Conf. on Simulation of Adaptive Behavior*, pp.460-468, 1993.
- [5] Craig W. Reynolds, "Flocks, Herds, and Schools: A Distributed Behavioral Model", *Proc. of SIGGRAPH*, pp.25-34, 1987
- [6] O. Miglino, H. H. Lund, and S. Nolfi, "Evolving Mobile Robots in Simulated and Real Environments", *Artificial Life*, vol.2, no.4, pp.417-434, MIT Press, 1996.
- [6] 민석기, 제갈욱, 강훈, "자율 이동 로봇군의 생명체 행동", 한국 퍼지 및 지능 시스템 학회(KFIS), 97년도 추계학술대회 논문집, vol.7, no.2, pp.83-86, 1997
- [7] J. A. Freeman and D. M. Skapura, *Neural Networks: Algorithms, Applications, and Programming Techniques*, Addison-Wesley, 1991
- [8] M. T. Hagan and H. B. Demuth, Mark Beale, *Neural Network Design*, PWS Publishing, 1996
- [9] J. S. R. Jang, C. T. Sun, and E. Mizutani, *Learning from Reinforcement, Neuro-Fuzzy and Soft Computing*, pp.258-300, Prentice Hall, 1997
- [10] R. S. Sutton, *Learning to Predict by the Methods of Temporal Differences*, *Machine Learning*, vol 8, pp9-44, 1988
- [11] C. J. C. H. Watkins, and P. Dayan, Technical Note : Q-Learning, *Machine Learning*, vol. 8, pp.279-292, 1992