Proceedings of KFIS Spring Conference 2005 Volume 15, Number 1

다수 로봇 제어를 위한 면적 기반 Q-learning

Area-Based Q-learning for Multiple Robots Control

윤한얼, 장인훈, 심귀보
중앙대학교 전자전기공학부

Han-Ul Yoon, In-Hoon Jang, and Kwee-Bo Sim
School of Electrical and Electronics Engineering, Chung-Ang University
E-mail: huyoon@wm.cau.ac.kr

ABSTRACT

본 논문에서는 다수개의 로봇을 효율적으로 제어하기 위한 면적기반 Q-learning에 대해 논한다. 각 로봇은 60°의 각을 이루도록 배치된 6개 센서를 가지고 있고 이를 통해 자신과 주변환경 사이의 거리를 센싱한다. 다음으로, 이 획득된 거리 데이터들로부터 6방향의 면적을 계산하여, 이후의 진행에 있어 보다 넓은 행동 반경을 보장해주는 영역으로 이동한다. 이 이동을 어떤 상태에서 다른 상태로의 전이로 간주, 이동 후 다시 6방향의 면적을 계산하여 이전 상태에서 현재 상태로의 행동에 대한 Q-Value를 업데이트 한다. 본 논문의 실험에서는 5개의 로봇을 이용해 장애물 사이에 숨어있는 물체를 찾아내는 것을 시도하였고, 3개의 서로 다른 제어 방법 - 랜덤 탐색, 면적 기반 탐색, 면적 기반 Q-learning 탐색 - 에 따른 결과를 나타내었다.

Key words : 면적 기반 행동 결정, Q-learning, 면적 기반 Q-learning

I. 서 론

최근들어 화재가 발생한 건물에서의 구조 활동이나 가스 누출 사고 지역의 정보 수집, 깊은 바다 속의 탐색 또는 극 지방과 같은 곳에서의 기후 조사와 같은 영역에서, 로봇은 사람들을 대신하여 작업을 수행하게 되었다. 때때로 다수의 소형 로봇들이 땅 아래 곤충의 집과 같이 사람이 직접 접근하기 힘든 영역으로 보내진다.

다수의 로봇을 보다 유연하고 강인하게 제어하기 위한 방법은 현재까지 많은 주목을 받아왔다. Parker는 다수 로봇의 작업 수행을 위해 휴리스틱 형태의 알고리즘을 제안하였다 [1]. Ogasawara는 다수의 로봇을 이용해 커다란 물체를 수송하기 위해 분산 로봇 제어 방식을 이용하였다 [2]. 본 논문에서는 다수의 로봇이 어떤 작업을 수행함에 있어 서로간의 충돌을 피하고, 자신만의 고유한 영역을 탐색하도록 하기위한 방법으로 면적 기반 행동 결정 (area-based action making) 방식을 제안한다. 이 면적 기반 행동 결정은 면적 기반 Q-Learning의 기초가 된다.

강화 학습은 agent로 하여금 주변 환경의 탐색을 통해 능동적으로 환경에 대한 행동을 결정하도록 한다. 보상값이 존재하는 어떤 불확실한 영역을 탐색하는 동안 agent는 연속적인 상태 공간을 따라 적절한 보상값을 전달함으로써, 임의의 상태에 대해 어떤한 행동을 취해야 할지를 학습하게 된다 [3]. 강화 학습을 구현하기 위한 많은 방법 중, 본 논문에서는 Q-learning을 이용하였다. 그 이유는 Q-learning은 불완전한 정보를 가진 Markovian 공간에서의 행동 결정에 대해, 어떤 상태와 행동으로 이루

어진 Q-함수를 기본으로 하여 문제의 해결에 쉬운 방법을 제공하기 때문이다 [4]. 또한 이 임의의 상태 공간을 실제로 물리적인 공간으로 간주될 수 있다. 본 논문에서는 면적 기반 탐색을 강화하기 위한 방법으로 면적 기반 Q-learning (area-based Q-learning)을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 면적 기반 행동에 대해 나타낸다. 3장에서는 면적 기반 Q-learning에 대해 논한다. 4장에서는 위의 방법들을 적용한 목표물 탐색의 실험 결과를 보이고, 5장에서는 결론 및 향후 과제에 대해 논한다.

II. 면적 기반 행동 결정 (Area-Based Action Making)

면적 기반 행동 결정 (area-based action making)은 로봇이 어떤 상태있을 때, 로봇의 다음 행동을 결정하기 위한 방법이다. 이 방법을 면적 기반 행동 결정이라 부르는 이유는 로봇이 자신 주변의 환경을 둘 사이의 거리가 아닌 자신 주변의 면적을 통해 다음 행동을 결정하기 때문이다. 이 방식의 핵심은 로봇으로 하여금 자신 주위의 불확실성을 점차적으로 줄여 나가도록 한다는 데 있다. 이 방법은 행동 기반 방향 전환 방식 (behavior-based direction change)과 많은 유사점을 가지고 있다 [5][6]. 면적 기반 행동 결정에서 로봇은 자신 주변 공간의 형태를 파악하고 보다 넓은 영역을 보장해주는 곳으로 이동하게 된다. 그림 1은 센서를 통해 단순히 거리 정보를 이용할 때와 면적 기반 방식을 이용할 때의 차이점을 보여 준다. 본 논문에서는 실험을 위해 자체 제작된 소형 로봇을 이용하였고, 6개의 센서를 60° 간격으로 로봇 주위에 배치하였다.

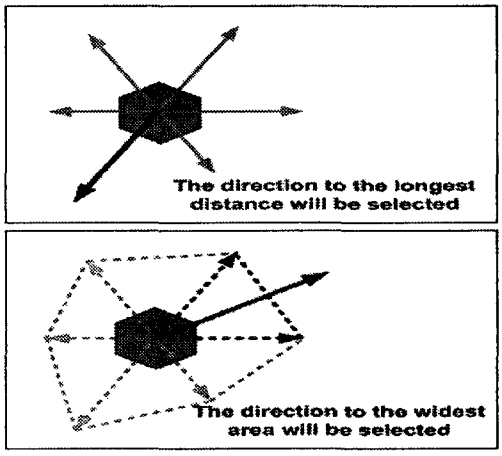

그림 1. 거리 기반 방식(위)과 면적 기반 방식(아래)에 의한 로봇의 행동 결정

그림 2 는 같은 환경에 있는 로봇들의 서로 다른 다음 행동 결정의 예를 통해, 면적 기반 행동 결정 방식의 장점을 보여준다. 그림 2에서 로봇은 4대의 장애물에 둘러싸여 있다. 왼쪽 그림과 같이 거리 방식을 따를 경우, 로봇은 남서쪽(+240°)에 장애물이 없다고 판단하여 계속해서 그 방향으로 진로를 결정할 것이다. 결국 로봇은 두 장애물 사이에서 벗어나기 어렵다. 반면에 오른쪽과 같이 면적 기반 방식을 이용할 경우, 로봇은 보다 넓은 영역을 보장해 주는 방향으로 행동을 취하므로 그림과 같이 로봇은 장애물에 둘러싸인 환경을 빠져나올 수 있다.

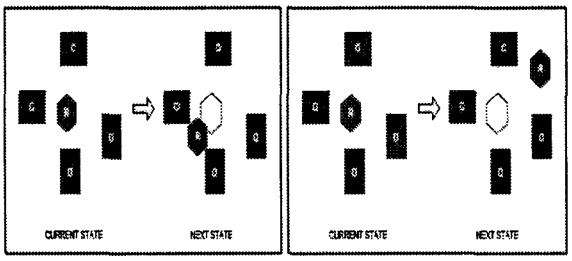

그림 2. 거리 기반 방식과 면적 기반 방식에 따른 행동 결정의 예

면적 기반 행동 결정 방법은 장애물 회피 뿐만 아니라, 각 로봇들로 하여금 자신의 고유한 영역을 탐색 할 수 있도록 해 준다. 2~3대의 로봇이 어떤 한 지점에서 서로 마주치게 되었을 경우를 생각해 보면, 각 로봇은 자신만의 보다 넓은 영역을 확보하기 위해 다음의 행동을 결정하게 된다. 따라서 각 로봇들은 각자만의 고유한 탐색공간을 따라 작업을 수행하게 된다 [7].

III. 면적 기반 Q-learning (Area-Based Q-learning)

Q-learning은 대표적인 강화 학습 알고리즘 중의 하나이다. Q-learning은 agent가 환경에 대한 선험적 정보를 가지고 있지 않을 때에도, 행동에 대한 보상값을 통해 최적의 행동 전략을 획득할 수 있도록 해준다 [8]. Q-learning 알고리즘을 표 1에 나타내었다. 여기서 s 는 상태를, a 는 행동을, r 는 보상값을, γ 는 Q-함수값의 조정을 위한 계수이다.

그림 3은 Q-learning에 대한 예를 보여준다. 각각의 정사각형은 상태를 나타낸다. 'R'은 로봇을 나타낸다. 상태의 천이에 따른 화살표 위에 나타난 값은 그 행동을 취함에 따른 Q 값을 나타낸다. 예를들어 초기 상태에서 오른쪽으로 상태를 천이하는데 따른 Q 값은 화살표 위의 $Q(s_1, a_{right}) = 72$ 와 같다.

표 1. Q-learning 알고리즘

For each s, a initialize the table entry $\hat{Q}(s,a)$ to zero
Observe the current state s
Do forever

- Select an action a and execute it
- Receive immediate reward r
- Observe the new state s'
- Update the table entry for $\hat{Q}(s,a)$ as follows:

$$\hat{Q}(s,a) \leftarrow r + \gamma \max_{a'} \hat{Q}(s',a') \qquad (1)$$

- $s \leftarrow s'$

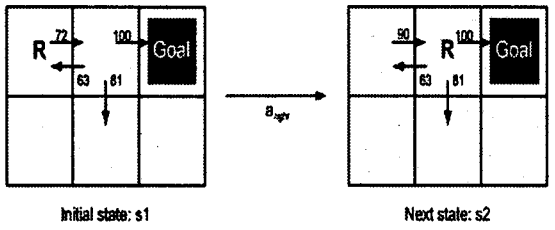

그림 3. Q-learning의 예

초기 상태에서 만약 로봇이 오른쪽으로 행동을 취한다면, 업데이트 되는 Q값은 $r=0$, $\gamma=0$ 이라 할 때

$$\begin{aligned}\hat{Q}(s,a) &\leftarrow r + \gamma \max_{a'} \hat{Q}(s',a') \\ &\leftarrow 0 + 0.9\max\{63,81,100\} \\ &\leftarrow 90\end{aligned} \qquad (2)$$

이 된다.

본 논문에서는 면적 기반 행동 결정을 강화하기 위해 Q-learning을 사용하였다. 로봇은 6개의 센서를 가지고 있으므로 기반이 되는 면적의 모양은 육각형이 된다. 따라서 어떤 임의의 상태의 로봇은 6방향으로 행동을 취할 수 있고, 6개의 Q값을 갖게된다. 그림 4에 Q-learning 적용의 간단한 예를 나타내었다. 만약 초기 상태에서 로봇이 +60° 방향으로 행동을 결정하였고, 상태 천이후 Area3가 가장 넓은 영역을 보장하였다고 가정하면, 이에 대한 Q값은 $\hat{Q}(s_1, a_{+60})$로 표현될 수 있고, 그 값은 0이 초기 상태에서의 보상값이라고 했을 때,

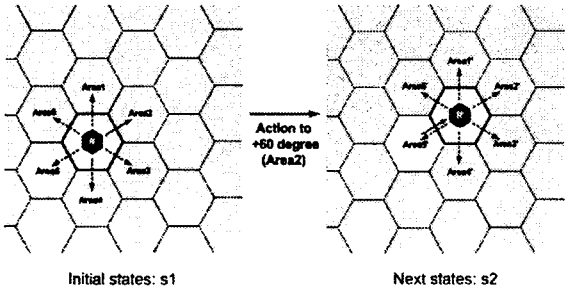

그림 4. 육각형 면적 기반 Q-learning의 예

$$\begin{aligned}\hat{Q}(s_1,a_{+60}) &\leftarrow r + \gamma \max_{a'} \hat{Q}(s_2,a') \\ &\leftarrow 0 + \gamma \max\{Area1, Area2, \cdots, Area6\} \\ &\leftarrow \gamma Area3\end{aligned} \qquad (3)$$

과 같이 된다. 초기 상태 이후의 보상값은 다음과 같이 결정된다.

$$r = \sum_{j=1}^{6} Area_j - \sum_{i=1}^{6} Area_i \qquad (4)$$

결과적으로 로봇은 이 Q값을 학습함으로써 진행 경로를 결정할 수 있다. 그러나 소프트웨어 상에서의 무한 반복은 현실 세계에서는 배터리 소비문제가 있으므로 구현이 불가능하다. 따라서 본 논문의 실험에서는 이전 상태에서 결정된 행동이 심각한 결과를 초래하였다면 다시 상태 천이 이전의 상태로 돌아가도록 해주었다. 전체 과정을 표 2에 나타내었다.

표 2. 육각형 면적 기반 Q-learning 알고리즘

For each s, a initialize the table entry $\hat{Q}(s,a)$ to zero

Calculate each 6-areas at the current state s

Do until task is completed.

- Take an action a to the widest area
- Receive immediate reward r
- Observe the new state s'

If $\hat{Q}(s',a')$ is greater or equal than $\hat{Q}(s,a)$

- Update the table entry for $\hat{Q}(s,a)$
- $s \leftarrow s'$

If $\hat{Q}(s',a')$ is too less than $\hat{Q}(s,a)$

- Move back to the previous state
- $s \leftarrow s$

IV. 다수 로봇에 의한 목표물 탐색 실험

본 논문의 실험에서는 5대의 로봇을 가지고 선험적 지식이 없는 환경에서 장애물 뒤에 숨어있는 목표물 탐색을 시도하였다. 목표물은 녹색의 로봇이며 특정 장애물 뒤에 정지상태로 숨어있다.

첫번째로, 랜덤 탐색 알고리즘을 적용하여 목표물 탐색을 시도하였다. 랜덤 탐색은 평균 이하의 성능을 가지므로 대부분의 경우 5대의 로봇 모두가 목표물 탐색에 실패하였다. 그림 5는 5대의 로봇이 랜덤 탐색을 이용하여 목표물 (흰색 화살표로 표시된 것)을 탐색하고 있는 과정을 보여준다.

그림 5. 랜덤 탐색에 의한 목표물 탐색

두번째로, 면적 기반 행동 결정 알고리즘을 적용한 경우, 5대의 로봇들은 보다 넓은 자신의 활동범위를 보장하는 곳으로 이동하게 되므로, 랜덤 탐색 보다 효율적으로 공간의 탐색을 수행할 수 있었다. 결과적으로 평균 2대 정도의 로봇이 목표물을 찾아내었다. 그림 6에서 검은색 화살표로 표시된 로봇들이 탐색에 성공한 로봇들이다.

그림 6. 면적 기반 행동 결정 알고리즘에 의한 목표물 탐색

마지막으로, 6-면적 기반 Q-learning 알고리즘을 적용하여 탐색을 수행하였다. 로봇들은 수행을 거듭할 수록 환경에 대한 정보를 학습해 가므로 10번의 수행을 통해 평균 2.8대 이상의 로봇이 목표물 탐색에 성공하였다. 그림 7은 탐색 과정을 보여준다.

그림 7. 면적 기반 Q-learning을 이용한 탐색 과정

V. 결론 및 향후과제

본 논문에서는 실제 5대의 로봇을 통한 선험적 지식이 없는 공간에서의 목표물 탐색을 위해 면적 기반 Q-learning 알고리즘을 제안하고 실험의 결과를 통해 이 알고리즘이 임의의 공간에서의 목표물 탐색에 새로운 방법이 될 수 있음을 보였다.

향후 그립퍼 (Gripper) 장착을 통해 다수 로봇에 의한 물체 수송, 대열을 갖춘 다수 로봇의 이동, object following 또는 path following 등의 로봇 기동에 관한 구현과 Fuzzy와 강화학습의 융합이나 TD (λ) 방법의 적용과 같은 심도 있는 알고리즘의 적용에 대한 연구가 뒤따라야 하겠다.

감사의 글: 본 연구는 과학기술부 뇌신경정보학 연구사업의 '뇌정보처리에 기반한 감각정보 융합 및 인간행위 모델 개발' 연구비 지원으로 수행되었습니다. 연구비 지원에 감사드립니다.

VI. 참고문헌

[1] L. Parker, "Adaptive action selection for cooperative agent teams," *Proc. of 2nd Int. Conf. on Simulation of Adaptive Behavior*, pp. 442-450, 1992.

[2] G. Ogasawara, et al., "Multiple movers using distributed, decision-theoretic control," *Proc. of Japan-USA Symp. On Flexible Automation*, vol. 1, pp. 623-630, 1992..

[3] D. Ballard, *An Introduction to Natural Computation*, MIT Press, Cambridge:1997.

[4] J. Jang, C. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computing, Prentice Hall, New Jersey:1997.

[5] W. Ashley, T. Balch, "Value-based observation with robot teams (VBORT) using probabilistic techniques," *Proc. of Int. Conf. on Advanced Robotics*, 2003.

[6] W. Ashley, T. Balch, "Value-based observation with robot teams (VBORT) for dynamic targets," *Proc. of Int. Conf. on Intelligent Robots and Systems*, 2003.

[7] P. Ogran, N. E. Leonard, "Obstacle avoidance in formation," *Proc. of IEEE Int. Conf. on Robotics and Automation*, vol. 2, pp. 2492-2497, 2003.

[8] T. Mitchell, *Machine Learning*, McGraw-Hill, Singapore:1997