Proceedings of KFIS Autumn Conference 2007 Vol. 17, No. 2.

로봇의 목표물 추적을 위한 SVM과 12각형 기반의 Q-learning 알고리즘

Dodecagon-based Q-learning Algorithm using SVM for Object Search of Robot

서상욱, 장인훈, 심귀보

Sang-Wook Seo, In-Hun Jang, and Kwee-Bo Sim

중앙대학교 전자전기공학부
(E-mail: kbsim@cau.ac.kr)

요 약

본 논문에서는 로봇의 목표물 추적을 위하여 SVM을 이용한 12각형 기반의 Q-learning 알고리즘을 제안한다. 제안한 알고리즘의 유효성을 보이기 위해 본 논문에서는 두 대의 로봇과 장애물 그리고 하나의 목표물로 정하고, 각각의 로봇이 숨겨진 목표물을 찾아내는 실험을 가정하여 무작위, DBAM과 AMAB의 융합 모델, 마지막으로는 본 논문에서 제안한 SVM과 12각형 기반의 Q-learning 알고리즘을 이용하여 실험을 수행하고, 이 3가지 방법을 비교하여 본 논문의 유효성을 검증하였다.

Key Words : SVM, Dodecagon-based Q-learning, DBAM, ABAM

1. 서 론

최근 들어 화재가 발생한 건물에서의 구조 활동, 가스 누출 사고 지역의 정보 수집, 깊은 바다 속의 탐색, 극지방과 같은 곳에서의 기후 조사와 같은 부분에서, 로봇이 사람을 대신하여 작업을 수행하고 있다. 특별히, 땅 아래 곤충의 집과 같은 사람이 직접 접근하기 힘든 곳의 탐색에서, 신뢰성 과 이용가치가 높은 정보의 획득을 위해 다수의 소형 로봇들이 보내진다. 다수의 로봇을 보다 유연하고 강인하게 제어하기 위한 방법은 현재까지 많은 주목을 받아왔다. Parker는 다수 로봇의 작업 수행을 위해 heuristic 형태의 알고리즘을 제안하였다[1]. Ogasawara는 다수의 로봇을 이용해 커다란 물체를 수송하기위해 자율 분산 로봇 제어 방식을 이용하였다[2]. 본 논문에서는 다수의 로봇이 어떤 작업을 수행함에 있어 서로간의 충돌을 피하고, 자신만의 고유한 영역을 탐색하도록 하기위한 방법으로 distance-based action making and area-based action making process의 융합 모델을 제안한다.

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 어떠한 행동을 취해야 할지를 학습하게 된다[3]. 강화 학습을 구현하기위한 많은 방법 중, 본 논문에서는 SVM을 바탕으로 한 Q-learning을 이용하였다. 그 이유는 Q-learning은 불완전한 정보를 가진 Markovian 공간에서의 행동 결정에 대해, 어떤 상태와 행동으로 이루어진 Q-함수를 기본으로 하여 문제의 해결에 쉬운 방법을 제공하기 때문이다[4]. 또한 이 임의의 상태 공간을 실제로 물리적인 공간으로 간주될 수 있다. 그리고 결과로 나타난 Q 값에 대하여 SVM을 사용함으로써 로봇의 행동에 대한 오차를 줄일 수 있었다[5]. 본 논문에서는 distance-based action making and area-based action making process의 융합 모델을 강화하기 위해 SVM과 12각형 기반의 Q-learning 알고리즘을 적용한다.

감사의 글 : 본 연구는 산업자원부의 2007년도 성장동력기술개발사업인 「집단 로봇 기술을 이용한 사회안전로봇 개발(세부과제: 로봇통제 및 환경기술개발)」에 의해 수행되었습니다. 연구비 지원에 감사드립니다.

본 논문의 2장에서는 distance-based action making and area-based action making process의 융합 모델에 대해 나타낸다. 3장에서는 SVM과 12각형 기반의 Q-learning 알고리즘에 대해 논한다. 4장에서는 위의 3가지 제어 알고리즘들을 적용한 목표물 탐색의 실험 결과를 보인다. 마지막으로 5장에서는 결론 및 향후 과제에 대해 논한다.

2. 로봇의 행동 결정 과정

2.1 DBAM and ABAM

Distance-based action making (DBAM) 과 Area-based action making (ABAM) 방법은 로봇이 다음 행동을 결정하는데 있어서 사용되는 방법이다. DBAM 방법은 로봇이 주위의 환경을 거리로 인식하는 방법으로 로봇과 물체 사이의 거리에 의해서 행동을 결정한다. 반면 ABAM 방법은 로봇이 자신 주변의 환경을 둘 사이의 거리가 아닌 자신 주변의 면적을 계산하여 얻어진 정보로부터 다음 행동을 결정하는 방법이다. 결국 ABAM 방법은 행동 기반 방향 전환(behavior-based direction change) 방식과 많은 유사점을 가지고 있다고 할 수 있다 [6][7]. 그림 1은 DBAM과 ABAM 방법이 행동을 선택하는 기준을 나타내고 있다.

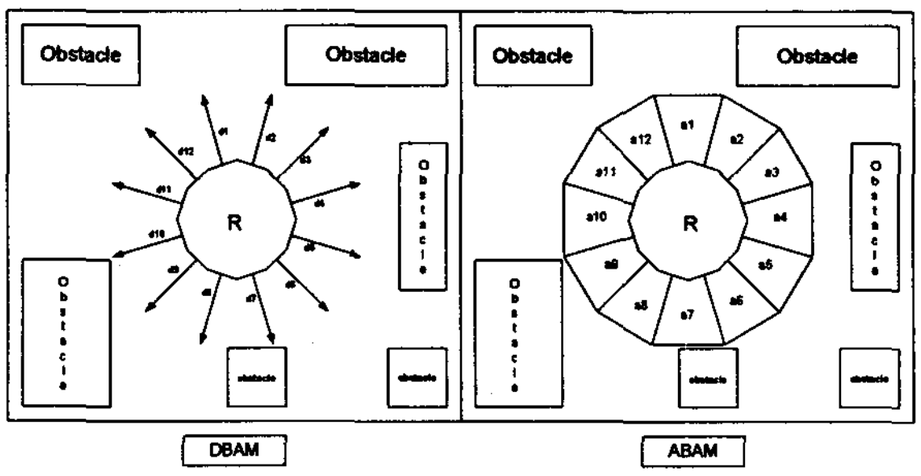

그림 1. DBAM과 ABAM의 행동 선택 기준

2.2 DBAM and ABAM의 융합모델

본 논문에서는 로봇이 다음 행동을 결정하는데 있어서 DBAM과 ABAM 방법을 융합한 모델을 사용한다. DBAM 방법을 통해서 로봇으로부터 가장 거리가 먼 방향을 선택하고, ABAM 방법으로 주위 환경에서 가장 넓이가 큰 공간을 선택하게 된다.

단순히 거리만으로 다음 행동을 결정하는 DBAM 방법은 계산양을 줄일 수 있다는 장점이 있지만, 올바르지 못한 행동을 선택할 확률이 높은 단점이 있다. 반명 행동을 선택할 때 단순히 넓이만을 고려하는 ABAM 방법은 올바른 행동을 선택할 확률이 있다는 장점이 있으나, 계산양이 많다는 단점이 있다.

DBAM과 ABAM 방법의 융합 모델은 행동을 선택할 때 거리가 가장 멀고, 넓이가 가장 큰 곳을 선택한다. 본 논문에서 선택한 방법은 로봇으로부터 거리가 가장 먼 방향을 선택하고, 선택되어진 방향 중에서도 가장 넓이가 큰 공간을 선택하는 방법이다. 그림 2는 DBAM과 ABAM 방법의 융합 모델에 관한 그림이다.

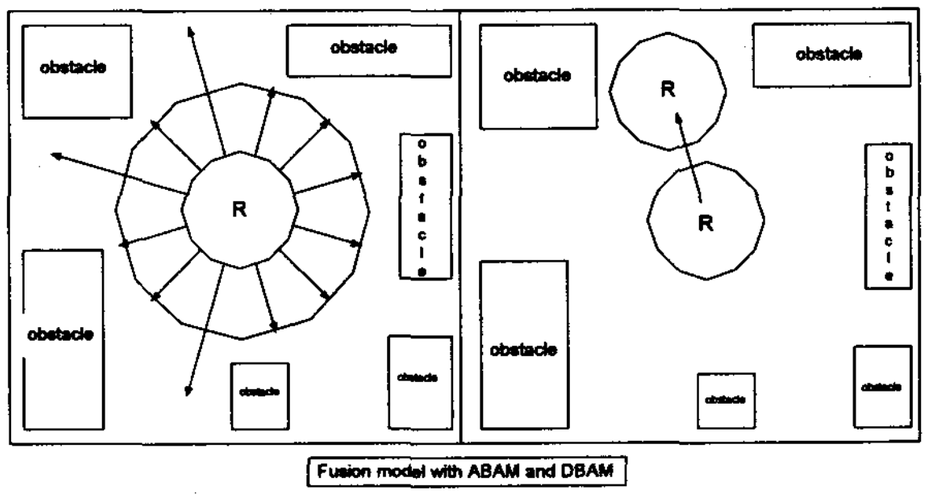

그림 2. DBAM과 ABAM의 융합 모델

3. SVM과 12각형 기반의 Q-learning 알고리즘

3.1 Q-learning

Q-learning은 강화학습으로 잘 알려진 알고리즘이다. 그리고 로봇이 효과적인 행동을 하기 위해서 보상의 개념을 이용해서 최적의 제어를 얻을 수 있다. 여기서 보상은 행동 후 보상을 하게 된다 [8][9]. Q-learning 알고리즘은 표 1에서 설명하는 것과 같다. 갱신될 Q 값은 다음의 식에 따라서 갱신된다.

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

표 1. Q-learning 알고리즘

For each s, a initialize table entry $\hat{Q}(s,a)$ zero observe the current state s.
Continue to infinity.
- Select action a and execute.
- Receive immediate reward r.
- Observe new state s'.
- Select action.
- Update table entry for $\hat{Q}(s,a)$.
- $s \leftarrow s'$.

3.2 12각형 기반 Q-learning 알고리즘

SVM을 이용한 12각형 기반 Q-learning 알고리즘은 12개의 초음파센서를 이용하여 12방

향으로 물체를 측정한다. 그리고 로봇 주위의 넓이를 12등분하여 인식한 후 각 넓이에서 물체의 넓이를 뺀 나머지 부분의 넓이와 물체와 로봇사이의 거리가 가장 긴 방향으로 행동을 취한다.

본 알고리즘은 기존의 Q-learning 알고리즘과 가장 큰 차이는 업데이트 방법이다. 기존의 Q-learning 알고리즘에서는 이전의 Q 값과 비교를 해서 최소화시킬 수 있는 방향으로 이전의 상태를 좋은 방향으로 학습해 나가는 방법이다. 하지만 본 연구에서 제안된 SVM과 12각형 기반의 Q-learning 알고리즘은 이전의 Q값에 의존하지 않고 그 상황에서 최적의 행동을 결정하게 된다. Q값의 갱신은 다음의 (2)식으로 이루어진다.

$$\begin{aligned}\hat{Q}(s,a) &\leftarrow r+\gamma \max_{s'}\hat{Q}(s',a) \\ &\leftarrow r+\gamma \max_{s'}\{(S_{area1}-S_{obstacle})l_{obstacle},(S_{area2}-S_{obstacle})l_{obstacle},\cdots, \\ &\qquad (S_{area12}-S_{obstacle})l_{obstacle}\}\end{aligned} \quad (2)$$

각각의 넓이가 동일하다고 했을 경우에는 로봇과 물체사이의 거리가 가장 먼 경우를 선택하게 되고, 만약 거리가 모두 동일하다고 하였을 경우에는 로봇 주위의 동일한 넓이에서 물체의 넓이를 뺀 공간 중에서 가장 넓은 공간을 로봇은 선택하게 되어있다.

1. $(S_{areaN}-S_{obstacle})$: largeness $l_{obstacle}$: largeness
2. $(S_{areaN}-S_{obstacle})$: largeness $l_{obstacle}$: smallness
3. $(S_{areaN}-S_{obstacle})$: smallness $l_{obstacle}$: largeness
4. $(S_{areaN}-S_{obstacle})$: smallness $l_{obstacle}$: smallness

여기서 4번째의 경우는 물체가 차지하는 넓이도 넓고 로봇과의 거리도 가깝기 때문에 선택되어질 확률이 가장 낮다.

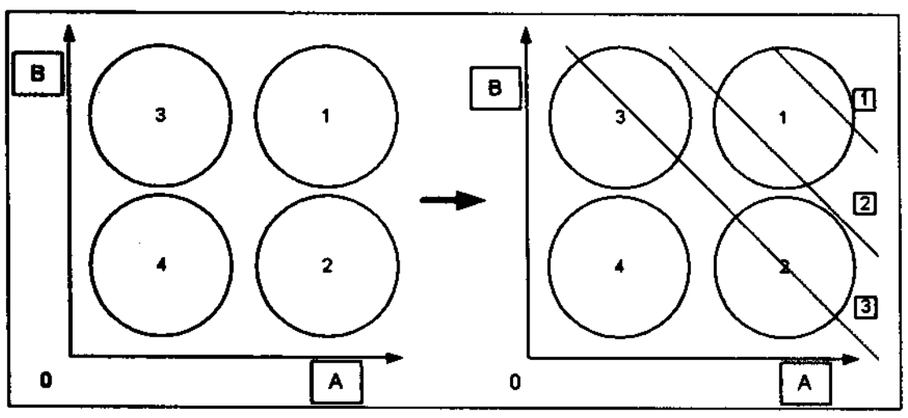

그림 3. 초평면에 의한 Q-값의 좌표

여기서 직선은 최적화 시킨 초평면이다. 다음의 (3)식은 초평면을 나타내는 식이다.

$$f(s)=\mathrm{sgn}(<S_{AreaN}-S_{obstacle}, l_{obstacle}>+b) \quad (3)$$

식에서 b는 임의의 상수이다. Q-값은 이러한 초평면에 의해서 상태값을 최적화 시키도록 학습해 나간다.

그림 4는 SVM과 12각형 기반의 Q-learning 알고리즘의 예이다. 로봇은 12개의 센서를 통해 주위환경을 인식하고 각 영역에서 로봇과 물체사이의 거리, 물체의 넓이를 뺀 넓이를 측정한다. 그 후 로봇은 최적인 공간으로 이동, 혹은 행동을 하게 된다.

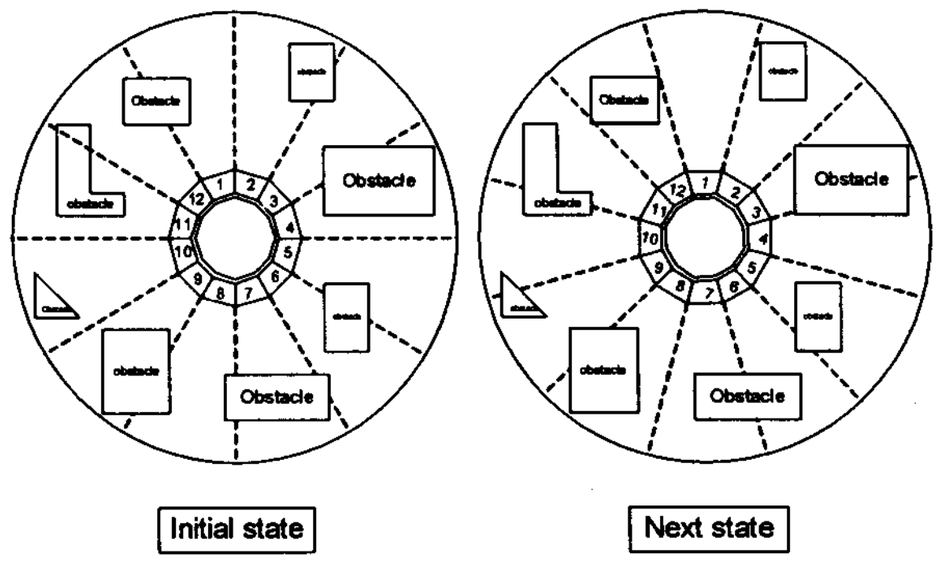

그림 4. SVM을 이용한 12각형 기반 Q-learning의 예

4. 시뮬레이션 및 결과

본 연구에서 제안한 알고리즘의 유효성을 보이기 위해서 본 논문에서는 3개의 알고리즘을 이용하여 모의실험을 수행하였고, 그 결과를 그림 5에 나타내었다. 총 6회의 탐색 시도에서, 랜덤 탐색의 경우 5번째 경우 1대의 로봇이 목표물을 찾아내었으나, 랜덤 탐색의 특성상 통계적인 의미를 부여하기는 어렵다. 다음으로 융합 모델인 경우, 모든 시행동안 평균적으로 1대 정도의 로봇이 목표물을 찾아내었다. 이것은 융합모델을 통해서도 탐색의 성능이 상당히 강화 될 수 있음을 나타낸다. 마지막으로 본 연구에서 제안한 SVM과 12각형 기반 Q-learning 알고리즘을 통한 탐색의 결과는 주목할 만한 결과이다. 총 6회 시행에 평균적으로 2대 정도의 로봇이 목표물 탐색에 성공하였다.

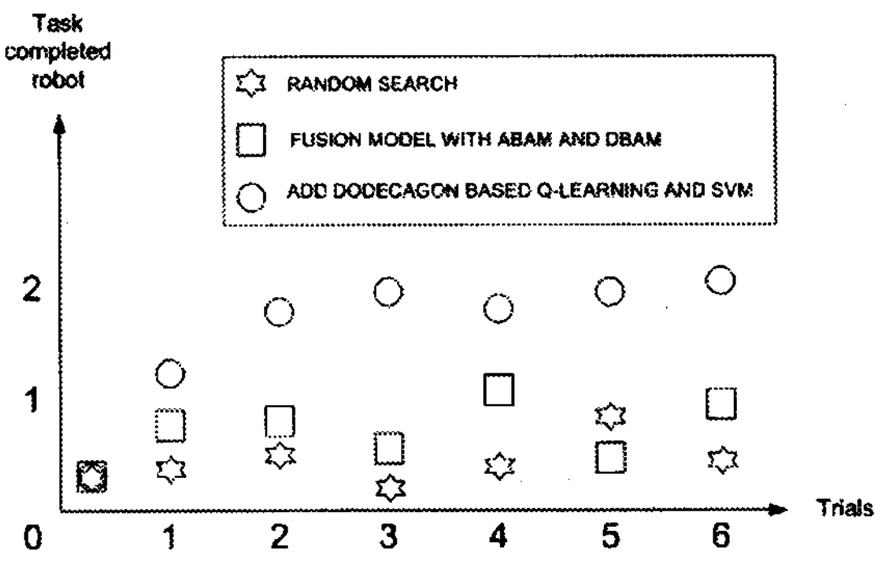

그림 5. 실험 결과

5. 결 론

본 논문에서는 먼저 선험적 지식이 없고 장애물이 놓여있는 공간에서의 목표물 탐색 알고리즘으로 융합 모델과 SVM과 12각형 기반의 Q-learning 알고리즘을 제안하였다. 또한 실제 2대의 로봇을 통해 위에서 언급한 조건의 환경에서의 목표물 탐색을 수행하고 그 결과를 보였다. 실험의 결과를 통해 이 알고리즘이 위와 같은 환경에서의 목표물 탐색에 새로운 방법이 될 수 있음을 알 수 있다.

향후 과제로는 첫째, 로봇들의 협조 행동 구현을 위해 목표물 발견 후 목표물에 접근하는 알고리즘의 구현이 필요하다. 둘째, 다수 로봇에 의한 물체 수송, 대열을 갖춘 다수 로봇의 이동, object following 또는 path following 등의 로봇 기동에 관한 구현과 Fuzzy와 강화학습의 융합이나 방법의 적용과 같은 심도 있는 알고리즘의 적용에 대한 연구가 뒤따라야 할 것이다. 마지막으로 본 논문에서 제안한 알고리즘은 향후 더 많은 로봇으로 이루어진 시스템에서 그 유효성을 좀 더 정밀하게 검증할 예정이다.

참 고 문 헌

[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, T. Omata, and T. Sato, "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*, The MIT Press Cambtidge, 1997.

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

[5] 이호근, 김명훈, 이지근, 정성태, "SVM-SMO와 Pan-Tilt 웹카메라를 이용한 실시간 얼굴 추적과 얼굴 인식," *한국정보과학회논문지*, 제31권, 제2호vol 31, no 2, pp. 679-681, 2004.

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

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

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

[9] C. Clausen and H. Wdchsler, "Quad-Q-learning," IEEE *Trans. on Neural Network*, vol. 11, pp. 279-294, 2000.