



An Action Select Reinforcement Learning Model in Multi Agent Environment

Bong-Keun Lee¹, Yoon-Ae Ahn²

¹*Dept. of Computer Engineering, Hannam University*

²*Dept. of Medical IT Engineering, Korea National University of Transportation*

A B S T R A C T

Reinforcement learning is concerned with how an agent ought to take actions in an environment so as to maximize some notion of long-term reward. In particular, a multi-agent system consisting of multiple interacting agents has increased state and action space as compared to a single agent system, so it must have an effective action selection mechanism. Reinforcement learning is used to evaluate and improve the effectiveness of a robotic soccer agent's action selection. That is, an agent that chooses its actions according to a certain action selection policy receives feedback regarding whether the chosen actions are desirable or not, and the agent learns to find optimal actions for various states in simulated soccer games based on the feedback. Possible states were identified by analyzing various situations/conditions arising in simulated soccer games, and then action selection policies were defined based on the analysis of a soccer agent's behavior patterns. In this paper can be exploited to develop optimized strategies and tactics for robot soccer systems, and it is also applicable to other multi-agent learning environments similar to the robot soccer game environment. In such environments, it is significant to acquire a policy that enables intelligent agents to work cooperatively to win the game.

© 2015 KKITS All rights reserved

KEYWORDS : Multi-agent, Reinforcement Learning, Optimal actions, ϵ -greedy, Robocup

ARTICLE INFO: Received 26 July 2015, Revised 14 August 2015, Accepted 14 August 2015.

*Corresponding author is with the Department of Medical IT Engineering, Korea National University of Transportation, 61 Daehak-ro, Jeungpyeong-gun Chungbuk

368-701, Republic of KOREA.
E-mail address: yeahn@ut.ac.kr

1. 서론

지능형 에이전트나 로봇의 발전은 인간이 행할 수 없는 다양한 일들을 해결해 줄 수 있는 분야에 적용되고 있다[1]. 기존의 단순한 작업만 수행하던 에이전트가 아닌 환경에 대한 효과적인 대처 능력을 가지며 스스로 문제를 해결할 수 있는 능력을 가진 로봇이나 에이전트가 필요하다. 에이전트는 주어진 환경에서 어느 정도 자율적으로 위임자를 대신하여 능동적으로 임무를 수행할 수 있는 지능형 프로그램으로 환경에 대한 효과적인 대처 능력이 중요하다. 에이전트의 환경에 대한 효율적인 대처 능력은 학습을 통해 이루어지며 지능형 에이전트의 학습을 위한 다양한 연구가 진행되었다[2][3].

로보컵(Robocup) 등장 이후 로봇 축구 시스템은 로봇의 주행제어, 비전 시스템, 멀티 에이전트 등 여러 분야의 연구 소재로 활용되어 왔으며 에이전트는 체스를 두는 프로그램으로부터 인간이 가기 힘든 미지의 인공 탐사선까지 매우 다양한 분야에 이용되었다[4][5]. 이와 같은 지능형 에이전트의 학습에 있어서 정해진 전략을 사용하는 교사 학습보다는 환경을 감지하여 스스로 인지하여 최적의 전략을 세울 수 있는 강화 학습과 같은 비 교사 학습이 더욱 효율적이다.

본 논문에서는 멀티 에이전트의 효율적인 학습을 위한 강화 학습 환경인 로보컵을 테스트 베드로 삼아 다양한 비교 실험을 전개하여 에이전트가 상황에 맞는 효율적인 행동을 학습 할 수 있도록 하였다. 경기 중의 다양한 상황들을 잘 표현할 수 있도록 세분화하여 상태를 나누고 패턴을 분석하여 행동을 정의하였다. 강화 학습을 통해 행동 선택을 결정하도록 하였으며 행동 선택 강화 학습을 통하여 에이전트가 최적의 정책에 빠르게 수렴할 수 있도록 학습 수렴속도를 향상 시켰다

본 논문의 구성은 다음과 같다. 2장에서는 지능

형 에이전트, 멀티 에이전트 환경과 강화학습의 이론적인 내용을 기술한다. 3장에서는 멀티 에이전트 환경에서의 행동 선택 강화 학습 모델을 제시한다. 4장에서는 강화 학습 모델의 학습 행동 정책의 효율성 평가를 위한 실험을 통해 제안된 모델의 성능을 평가한다. 5장에서는 본 연구결과의 활용에 대해 논하고 결론을 맺는다.

2. 멀티 에이전트와 강화학습

2.1 멀티 에이전트

에이전트는 인공지능 분야에서 다양한 방법으로 연구되고 있는 개념으로 사실상 인공지능 연구의 최종 목표가 사람과 유사한 지적 능력을 소유한 에이전트의 개발이라고 할 수 있다. 지능형 에이전트는 자율 행동과 다른 에이전트와의 상호작용 및 환경 변화에 대한 적응과 같은 지적인 능력을 갖추고 있어 새로운 분야에 활용이 가능하다. 에이전트들은 개별적으로 구성되고 유지될 수 있으며 지식, 추론, 계획 능력과 환경 변화에 대한 적응, 학습, 인식능력 등 자율로봇과 관련된 특징을 가질 수 있다.

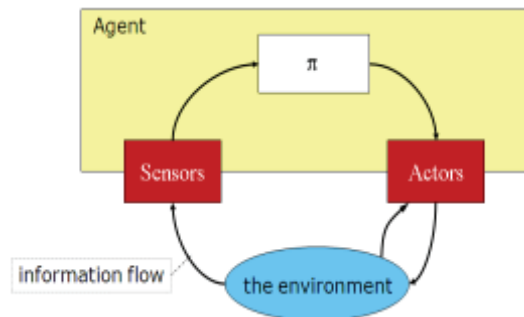


그림 1. 에이전트 동작 과정
Figure 1. Agent Acting Process

에이전트의 자율적인 행동은 환경과의 상호 작용으로부터 나온다. <그림 1>에서 보는 바와 같이 에이전트가 환경의 상태를 감지하고 행동을 선택하여 수행 하면, 환경은 그 행동의 영향으로 변화가 있게 되고 에이전트는 환경의 변화된 상태를 다음 상태로 다시 감지하는 과정을 반복하게 된다.

에이전트가 가진 특성들로는 상황에 적절하게 환경의 변화에 응답하는 반응성과 자신의 행동을 스스로 제어해 나갈 수 있는 자율성, 시간에 대하여 연속적으로 행동하는 연속성 및 에이전트의 행동이 경험에 따라서 변화하는 적응성 등의 특성이 존재한다. 그 이외에 환경변화에 대해 단순히 행동하지 않고, 사전 행동과 목적을 갖고 대처하는 목적 지향성과 스스로 움직일 수 있는 이동성, 유연성, 감성 등의 다양한 특성을 지니고 있다

멀티 에이전트 분야에서의 주요한 쟁점은 상호 협력과 경쟁을 통해 작업을 수행함으로써 최종의 목표를 달성하는 것이다[6]. 경쟁적 관계의 에이전트들의 작업이라면 각각의 에이전트는 서로의 이익만을 위해 최선의 정책을 수행한다. 멀티 에이전트 시스템의 구성에 사용되는 각각의 에이전트들은 에이전트가 원하는 목표, 환경에 관한 지식, 가능한 행동에 대해서 같은 구조를 갖는다. 또한 각 에이전트들이 행동을 선택하는 과정도 같은 절차를 행한다. 그러나 각 에이전트들 사이에 다른 점은 각 에이전트가 위치한 상태에서의 입력에 따라서 다른 행동을 한다. 따라서 각 에이전트가 위치한 상태에 따라서 각 에이전트들이 해야 할 행동은 서로 다르다.

멀티 에이전트 환경은 <그림 2>에서 보는 바와 같이 각각의 에이전트가 특정 도메인에 한정적인 작업을 수행하는 단순 에이전트의 형태를 가지거나 다수의 에이전트들이 상호작용과 협력을 통

해 복잡한 작업을 수행한다. 이러한 협동과 경쟁의 문제를 해결하기 위해서는 학습이 필요하며 학습은 단순히 사물을 인식하는 것이 아닌 새로운 정보를 통하여 얻어진 지식을 수정 할 수 있어야 하고 정확한 지식으로 개선해 나가는 능력을 갖도록 하는 것이다. 일반적으로 멀티 에이전트 학습에서는 교사 학습처럼 예제가 주어지지 않고 성취해야 할 목표와 행동을 평가하는 보상 함수가 주어진다.

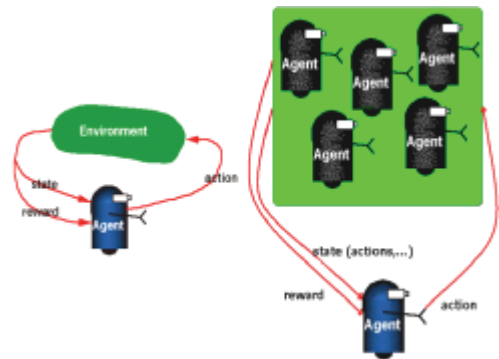


그림 2. 멀티 에이전트 환경
Figure 2. Multi-Agent Environment

2.2 강화 학습

Minsky에 의해 소개된 강화 학습은 심리학 분야에서 동물의 학습을 연구하는 과정에서 기원하였다 [7]. 강화 학습은 동적 프로그래밍과 교사 학습을 혼합한 형태의 학습 방법으로서 학습을 수행하는 에이전트는 에이전트의 외부에 존재하는 환경과 시행-착오를 통해 상호 작용하면서 학습한다. 보상이란 개체가 수행한 행동에 대한 환경에서 주는 잘하고 못함을 나타내는 값이다. 일반적으로 강화 학습을 수행하는 에이전트와 환경과의 관계는 <그림 3>과 같이 구성되어 있다.

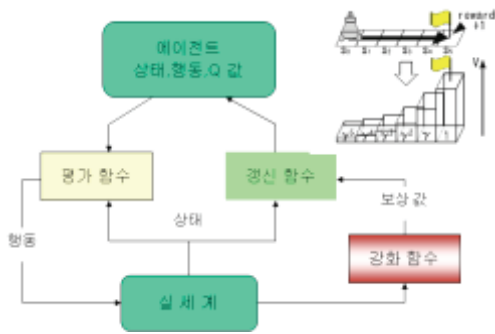


그림 3. 강화 학습 과정
Figure 3. Reinforcement Learning Process

강화 학습은 정확한 답이 그대로 주어지는 것이 아니라 선택한 행동의 좋고 나쁨에 따라 주어지는 보상을 이용하여 보다 좋은 결과를 얻을 수 있는 행동을 학습한다. 특히 단일 에이전트에 비해 상태 공간과 행동공간이 매우 커지는 멀티 에이전트 시스템인 경우 효과적인 강화 학습을 위해서는 적절한 행동 선택 전략이 마련되어야 한다. 멀티 에이전트 시스템을 구현할 때에는 여러 수의 개체들로 이루어진 지능 시스템의 구현 원리를 정하고 그 안에서 에이전트들이 독자적인 행동을 하면서 그와 동시에 서로 협력적인 행동도 할 수 있도록 하는 동작 메커니즘을 제공해 줄 수 있어야 한다.

동적 환경과 시행착오를 통해 상호작용하면서 학습을 수행하는 강화 학습 시스템은 학습을 하는 에이전트와 외부 환경과의 인터페이스, 에이전트의 학습목표, 외부 환경으로부터 주어지는 강화 값, 주어진 상태가 얼마나 좋은지 또는 주어진 상태에서 에이전트가 수행한 행동이 얼마나 좋은가를 평가하는 함수, 그리고 강화 학습의 최종 목표라 할 수 있는 최적 값으로 구성된다.

강화 학습의 목적은 상태와 행동을 사상시키는 최적의 정책(optimal policy)을 구성하는 것이다. 강화 학습을 이용한 사례로는 자율 이동 로봇의 주행 알고리즘, 생체 모방 시스템 및 인공 생명체의

지능 알고리즘, 게임 이론에서의 해석 알고리즘, 로봇 제어 등에 강화 학습 방법을 이용하는 연구가 진행되고 있다[6].

강화 학습의 가장 큰 특징은 에이전트의 학습목표를 정형화(formalize)하기 위해 강화 값을 사용하는 것이다. 즉, 에이전트의 학습목표는 학습과정에서 각 시간 단계마다 외부 환경으로부터 에이전트에게 전해지는 강화 값들의 합이 최대가 되도록 하는 것이다.

Q-학습은 각 상태와 행동의 쌍에 대한 Q-함수 값을 학습하는 강화 학습의 하나이며 환경에 대한 사전 모델이 필요하지 않은 학습방법이다[6]. Q-학습의 가장 큰 특징은 환경모델, 즉 보상함수 및 상태 결정함수가 필요 없는 강화 학습이라는 점이다. Q-학습에서는 상태-행동에 대한 평가함수 $Q(s, a)$ 값을 예측하고 이 Q값을 기초로 행동을 결정한다. 그러나 사전에 정확한 평가함수 $Q(s, a)$ 값을 알 수 없기 때문에 에이전트는 경험을 통해 점진적으로 최적의 Q함수 값을 찾아간다. 최적의 Q-함수 값에 수렴하기 위해서는 가능한 모든 상태와 행동에 대한 충분한 반복경험이 필요하다. 따라서 Q-학습 과정에서는 모든 상태-행동 쌍에 대한 평가함수 $Q(s, a)$ 값을 저장 운영할 Q-테이블이 필요하다.

표 1. Q-학습 테이블
Table 1. Q-Learning Table

State/Action	Action-1	Action-2	...	Action-m
State-1	Q11	Q12	...	Q1m
State-2	Q21	Q22	...	Q2m
:	:	:	:	
:	:	:	:	
State-n	Qn1	Qn2	...	Qnm

3. 멀티 에이전트 행동선택 강화학습

Q-학습의 수렴은 모든 행동을 충분한 횟수로 선택할 수 있는 경우에는 행동 선택 방법에 무관하게 성립한다. 따라서 개체에 의한 행동의 선택 방법은 무작위 선택이라도 크게 관계가 없다. 대부분의 강화 학습에서는 아직 Q-값 수렴하고 있지 않은 학습과정에 있어서도 가능한 한 대부분의 보상을 받는 행동을 선택하도록 요구하는 경우가 많다. 따라서 학습에 의하여 행동을 선택하는 경우, 이후에 주어지는 보상을 개선해 나가는 행동을 선택하도록 하는 탐사전략으로는 ϵ -greedy 선택과 같은 방법들이 있다.

ϵ -greedy 선택은 semi-uniform policy라고도 하며, ϵ 의 확률로 임의의 행동값 함수를 갖도록 하며, 그 이외의 경우에는 최대의 행동값 함수를 가지는 행동을 선택하도록 하는 방법이다.

Q-학습은 Q-함수의 값을 기준으로 상태에 대한 최적의 행동을 선택한다. 만일 현재의 Q-함수에 의하여 항상 최적의 행동만을 선택한다면 어떤 행동들은 계속해서 선택되지 않는 경우가 발생한다. 모든 행동을 경험하지 않는다면 에이전트는 최적의 행동을 발견할 수 없다. 환경의 정보를 얻기 위해서는 탐험 전략에 따라서 최적이지 아닌 행동들도 선택해야 한다. 그러나 탐험에 치우치다 보면 학습속도가 느려진다. 그러므로 탐험과 Q-함수 이용의 조율이 필요하다. 에이전트가 환경에 대한 정보를 구하기 위해서 탐험 전략에 의하여 행동을 선택하는 것을 탐험(exploration)이라 한다. 에이전트가 현재까지 학습한 지식을 바탕으로 최적의 행동을 선택하는 것을 지식의 이용(exploitation)이라 한다.

탐험과 지식 이용의 조율 문제를 해결하는 몇 가지 방법이 있다. 첫째, 모든 행동을 시도한 후, Q값에 따라서 행동을 선택하는 방법이 있다. 가장 간단한 방법이다. 둘째, 시간의 일정한 비율만큼

임의의 행동을 선택하는 방법인 ϵ -greedy 전략이다. 파라미터 ϵ 는 임의로 행동을 선택하는 시간 비율을 조정한다. ϵ 값이 1에 가까울수록 탐험하는 시간이 많아지고 0에 가까워지면 현재까지 학습한 지식에 의하여 행동을 선택하게 된다.

행동 선택 강화 학습 모델에서는 최적의 행동을 학습하기 위하여 <그림 4>에서와 같이 에이전트가 최종 목표값에 도달하는 최대 Q값을 갖도록 학습하며 ϵ -greedy 선택을 통한 최적의 행동 선택을 학습한다.

학습에서 임의의 초기 상태를 선택하고 행동을 취하면서 목표에 도달할 때까지 과정을 하나의 에피소드라고 정의한다. Episodic Q-학습은 이러한 에피소드를 여러 상태와 행동 쌍들의 순차적 집합이라 보고 에피소드의 모든 상태와 행동 쌍 순차의 역순으로 Q-테이블의 Q-함수를 갱신하는 학습법이다. Episodic Q-학습은 관련된 에피소드를 사용해 Q-함수를 갱신하는 학습법이기 때문에 여러 에이전트들이 에피소드를 공유할 수 있는 다중 에이전트 환경에 적합한 학습법이다.

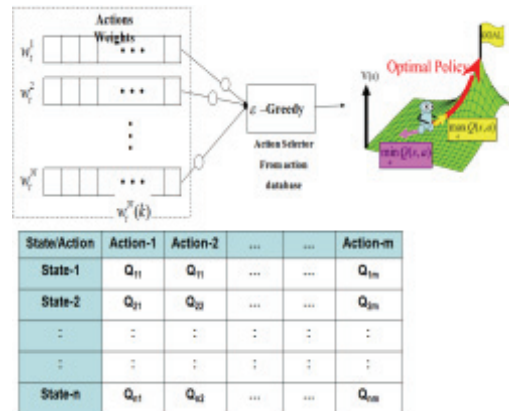


그림 4. 행동 선택 강화 학습 모델
Figure 4. Action Select Reinforcement Model

효율적인 로봇축구 시스템을 위해 경기 중 발생

되는 수많은 상황들을 최대한 고려할 수 있도록 여러 조건들을 이용하여 다양하게 상태를 정의하였다. 또한 에이전트가 정의된 행동 정책에 따라 동작할 수 있는 각각의 행동 함수를 만들었다. 다음으로 순수하게 에이전트의 행동만의 성능을 평가할 수 있도록 하기 위해 간소화된 실험 환경인 제한된 에피소드 환경을 구성하고, 여러 번의 실제 경기를 통해서 각 상태와 행동마다 Q값을 가지는 Q테이블을 작성할 수 있었다. Episodic Q-학습을 사용함으로써 기존의 Q-학습에 비해 Q-테이블이 신속하게 채워지므로 학습속도도 빨라지게 되었다.

4. 실험 및 평가

제안한 강화 학습 모델의 학습 행동 정책의 효율성 평가를 위하여 keeper의 정책을 Always Hold, Random, Hand-coded, Learned로 구분 하여 비교 실험하였다. 행동 정책의 효율성을 평가는 keeper의 평균 볼 점유 시간으로 측정하였다.



그림 5. Keep away 실험 화면
Figure 5. Keep away Main Screen

강화 학습 환경인 로보컵 Keep away를 테스트 베드로삼아 멀티 에이전트 행동 선택의 효율성을

확인하였다.

Keep away는 로보컵 축구 시뮬레이션 게임을 기반으로 강화 학습 테스트용으로 만들어진 도구이며 로보컵 축구 서버로부터 Keep away에 필요한 정보를 획득하여 시뮬레이션 한다. Keep away는 <그림 5>에서 보는 바와 같이 3명의 keeper와 2명의 taker로 구성되며 로보컵 축구 시뮬레이션 리그의 제한된 영역을 사용한다.

keeper의 목표는 최장시간 동안 공을 점유하거나 유지하는 것이고 taker의 목표는 공을 뺏는데 있다. 파라미터는 영역의 크기, keeper와 taker의 개수이다. Keep away는 에피소드 방식으로 이루어지며 keeper가 공을 빼앗기거나 영역 밖으로 나갔을 때가 하나의 에피소드가 된다. 각각의 에이전트는 공과 상대편 에이전트 등의 정보를 받아들이며 가장 기본적인 행동인 kick, turn, dash를 수행한다.

```

# Keepers: 3
# Takers: 2
# Region: 20 x 20
# Description of fields:
# 1) Episode number
# 2) Start time in simulator steps (100ms)
# 3) End time in simulator steps (100ms)
# 4) Duration in simulator steps (100ms)
# 5) (out of bounds / (taker away)
#
# 0 2 2 0
# 2 42 40 1
# 3 42 102 60 1
# 4 102 139 37 1
# 5 139 178 39 1
# 6 178 239 61 1
# 7 239 308 61 0
# 8 308 372 71 1
# 9 372 428 69 1
# 10 428 483 57 0
    
```

그림 6. 3Vs.2 keep away 실험 테이블
Figure 6. 3Vs.2 keep away Simulation Table

에피소드를 단위로 keeper의 볼 점유 시간이 얼마나 지속되는가를 학습의 측정치로 사용하였으며 총 50시간씩 10회에 걸쳐 실험 하였으며 세 가지 경우로 구분하여 비교 실험하였다.

첫 번째로 플레이어를 제한된 도메인에서 <그림

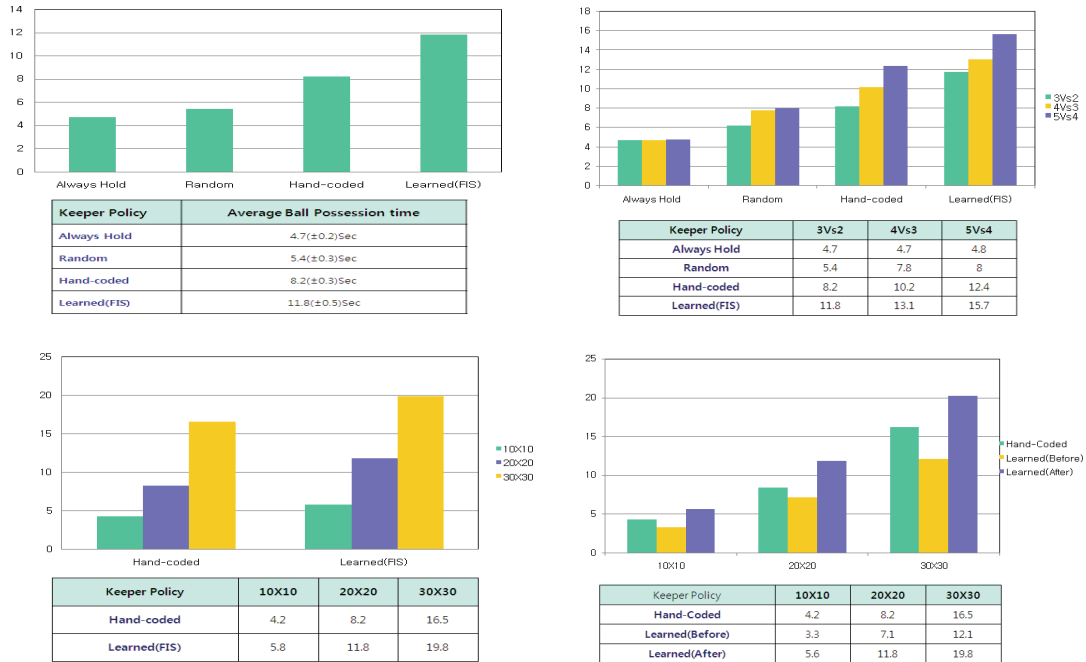


그림 7. Keeper의 평균 볼 점유 시간
Figure 7. Keeper's Average Ball Possession Time

6>과 같이 3Vs.2, 4Vs.3, 5Vs.4로 확장해가면서 경기를 진행하여 keeper의 볼 점유시간을 측정하였다. 두 번째로 도메인을 기준으로 10X10, 20X20, 30X30(m)으로 확장해가면서 점유시간을 측정하였다. 마지막으로 강화 학습 모델을 적용하기 전과 후의 성능 비교를 하기 위하여 학습하기 전과 후를 비교 실험하였다.

<그림 7>의 keeper의 행동정책 별 볼 점유시간 비교 실험 결과에서 keeper의 행동정책은 제안모델을 적용한 정책을 사용한 경우 다른 세 가지 정책에 비해 볼 점유시간이 향상됨을 보이고 있다. 따라서 강화 학습에 있어 퍼지 추론을 이용한 사전지식을 적용함으로써 에이전트의 행동선택의 효율성을 높일 수 있음을 확인하였다.

5. 결론

에이전트는 환경으로부터 최적의 행동을 스스로 찾아낼 수 있어야 하며 상황에 맞는 적절한 행동 선택이 필요하다. 기존의 고정된 정책을 사용하는 경우 에이전트의 행동이 정해진 정책에 의해 결정되기 때문에 최적의 정책을 학습하기 어렵다. 또한 다양한 상황에 맞는 행동 선택에 대한 평가의 부족으로 인해 행동 선택의 효율성을 기대하기 어렵다. 본 논문에서는 로봇 축구경기에서 에이전트의 효율적인 경기 수행을 위해 멀티 에이전트 환경에서의 행동 선택 강화 학습 모델을 제안하였다. 특정 상황에서 에이전트가 스스로 최적의 전략을 찾을 수 있도록 하였으며 제안된 모델을 활용하여 멀티 에이전트의 환경에서 에이전트의 행동 정책

의 효율성 평가와 로봇축구 시스템의 전략 및 전술에 활용이 가능하다.

References

- [1] Matthew E. Taylor, and Peter Stone, *Representation transfer for reinforcement learning*, AAAI 2007 Fall Symposium, Arlington, Virginia, 2007.
- [2] Samuel Barrett, and Peter Stone, *Cooperating with unknown teammates in complex domains: A robot soccer case study of Ad hoc teamwork*, In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 2015.
- [3] Yang, E., and Gu., *Multi-agent reinforcement learning for multi-robot system: A survey*, University of Essex Technical Report CSM-404, 2004.
- [4] Piyush Khandelwal, and Peter Stone, *Multi-robot human guidance using topological graphs*, In AAAI Spring 2014 Symposium on Qualitative Representations for Robots (AAAI-SSS), March 2014.
- [5] Alexander A. Sherstov, and Peter Stone, *Function approximation via tile coding: Automating parameter choice*, In Proc. Symposium on Abstraction, Reinforcement, and Approximation(SARA-05), 2005.
- [6] Noa Agmon, and Peter Stone. *Leading Ad hoc agents in joint action settings with multiple teammates*, In Proc. of 11th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS), June 2012.
- [7] Sutton, R.S., and Barto, A.G., *Reinforcement learning: An introduction*, MIT Press, 1998.
- [8] Fangkai Yang, Piyush Khandelwal, Matteo Leonetti, and Peter Stone. *Planning in answer set programming while learning action costs for mobile robots*. In AAAI Spring 2014 Symposium on Knowledge Representation and Reasoning in Robotics (AAAI-SSS), March 2014.
- [9] Whiteson, S., and Stone, P., *Concurrent layered learning*, Second International Joint Conference on Autonomous Agents and Multiagent Systems. pp. 193-200. {ACM } Press, New York, NY (2003).
- [10] David McAllester, and Peter Stone, *Keeping the ball from CMUnited-99*, Robocup-2000: Robocup IV, Springer Verlag, Berlin, 2001.
- [11] N. Hansen. *The CMA evolution strategy: A Tutorial*, January 2009.
<http://www.lri.fr/~hansen/cmatutorial.pdf>.
- [12] Patrick MacAlpine, Katie Genter, Samuel Barrett, and Peter Stone, *The RoboCup 2013 drop-in player challenges: Experiments in Ad hoc teamwork*. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 2014.

멀티 에이전트 환경에서의 행동 선택 강화 학습 모델

이봉근¹, 안윤애²

¹한남대학교 컴퓨터공학과

²한국교통대학교 의료IT공학과

요 약

강화 학습은 에이전트가 주어진 환경과 상호작용을 통하여 환경으로부터 받는 보상 값의 합을 최대화 할

수 있는 행동정책을 찾아야 하는 최적화 문제이다. 특히, 에이전트의 환경에 대한 대처 능력은 매우 중요한 요소로 에이전트가 현재 상황을 인식하여 상황에 맞는 적절한 행동을 선택할 수 있는 행동선택 전략이 요구된다. 멀티 에이전트 환경에서 에이전트의 행동 선택의 효율성을 평가하기 위하여 행동 선택 강화 학습을 통해 학습 행동을 평가하였다. 평가된 행동을 기반으로 효율적인 행동 선택을 학습하였다. 효율적인 로봇축구 시스템을 위해 경기 중 발생하는 수많은 상황들을 최대한 고려할 수 있도록 여러 조건들을 이용하여 다양하게 상태를 정의하였으며 로봇 축구 시스템에서 다양한 상황에 맞는 행동 선택 모델을 제안하였다. 경기 중의 다양한 상황을 세분화하여 상태를 분리하고 에이전트의 패턴 분석을 통한 행동 정책을 정의하였고 이를 기반으로 최적의 행동에 수렴할 수 있게 하여 로봇 축구 경기의 전략 및 전술에 활용 가능한 모델을 개발하였다. 논문에서 제안한 멀티 에이전트 행동 선택 강화학습 모델은 지능형 멀티 에이전트 학습에서 행동 선택의 효율성 평가와 로봇 축구 시스템의 전략 및 전술에 적용 가능하며 로봇 축구 시스템과 유사한 환경의 멀티 에이전트 학습에 활용될 수 있다.



Yoon-Ae Ahn received the MS degree in computer science from Chungbuk National University, Korea in 1996 respectively and also received Ph.D. in computer science from Chungbuk

National University, Korea in 2003. Since then, she has been with the dept. of Medical IT Engineering, Korea National University of Transportation. Her main research interests include mobile applications and medical information systems.

E-mail address: yeahn@ut.ac.kr



Bong-Keun Lee received the bachelor's degree in the Department of Computer Engineering from the Han-Nam University in 1997. He received the M.S. degree

the Department of Computer Engineering from the Han-Nam University in 1999 and the Ph.D. degree in the Department of Computer Science from Chungbuk National University in 2010. His current research interests include artificial intelligence, intelligent systems, Information Security, Mobile & Ubiquitous Web Service platform, Sensor Database. He is a life member of the KIPS.

E-mail address: bong9065@daum.net