

분류자 시스템을 이용한 축구 로봇의 행동 전략

Behavior strategies of Soccer Robot using Classifier System

김지윤, 이동욱, 심재윤, 심귀보
중앙대학교 전자전기공학부

Jee-Youn Kim, Dong-Wook Lee, Jae-Yoon Sim, and Kwee-Bo Sim
School of Electrical and Electronic Engineering, Chung-Ang University
E-mail : kbsim@cau.ac.kr

ABSTRACT

분류자 시스템은 유전자 알고리즘(Genetic Algorithm : GA)을 이용하여 새로운 규칙 집합을 발견하는 시스템이다. 또 로봇 축구 시뮬레이션 게임(SimuroSot)은 시간에 따라 상태가 변화하는 동적인 시스템이다. 본 논문에서는 GBML(Genetic Based Machine Learning)의 한 갈래이자 미시간 접근 방법을 기반으로 하는 Zeroth Level Classifier System(ZCS)을 SimuroSot에 적용하여 게임 전략을 구성하는 새로운 규칙의 발견과 학습에 의한 축구 로봇의 행동전략 알고리즘을 제안하고, 시뮬레이션을 통하여 본 전략의 유용성을 확인한다.

Key words : 로봇 축구, 분류자 시스템, 유전자 알고리즘, 기계학습, 자율이동로봇

I. 서 론

로봇 축구 시스템은 동일한 구조를 가진 여러 대의 로봇이 서로 역할을 분담하고 협조하면서 일정 시간동안 서로 경쟁하여 더 많은 득점을 한 팀이 승리하도록 구성된 시스템이다.

또, Holland에 의해서 제안된 GBML 즉 분류자 시스템은 내부의 적합도를 변화시키는 방법으로 강화학습 기법을 사용하고 유용한 새로운 일반적 규칙을 찾기 위해 유전자 알고리즘을 사용하는 parallel production rule system이다[1-2]. 그러나 이 시스템은 동적으로 변화하는 외부 환경의 상태를 모델링 할 때 각 분류자들의 조건부와 행동부가 같은 경우의 수를 가지도록 설계되어야 하며 활성화 된 분류자 행동부의 메시지가 다른 분류자를 활성화시키는 역할을 하는 메시지인지 아니면 효과기를 통하여 외부 환경으로 내보내어지는 메시지인지를 적절하게 결정하여야 하는 등의 몇 가지

어려운 문제를 해결해야 한다. 이런 문제들을 해결하기 위해 Wilson은 Holland의 분류자 시스템의 구조를 간략화 시킨 Zeroth Level Classifier System(ZCS)을 제안했다[3]. Wilson은 전체적인 구조는 Holland의 분류자 시스템을 받아들였으나 알고리즘을 간략화 하여 좀더 이해하기가 쉽고 문제에 대한 모델링을 수월하도록 하였다.

본 논문에서는 가상 환경으로써 로봇 축구 시뮬레이터를 사용하고 ZCS을 이용하여 동적으로 변화하는 게임 상태 속에서 상대팀의 전략에 대해 새로운 유용한 규칙을 발견하여 학습하는 잠재적인 적응 능력을 가진 전략 생성 시스템을 구현한다. 그리고 사람에 의해 임의로 구성된 전략을 가진 상대팀과의 경기 결과를 분석함으로써 ZCS의 유용성을 확인한다.

II. 로봇 축구 시뮬레이션을 위한 ZCS

2.1 ZCS의 구현

그림 1은 로봇 축구 시뮬레이션에 사용된 ZCS의 구조를 나타낸 그림이다. 여기서 외부 환경(Environment)은 로봇 축구 시뮬레이션

본 연구는 한국과학재단 목적기초연구 (과제번호 : 2000-2000-2-30300-003-3) 지원으로 수행되었습니다. 연구비 지원에 감사드립니다.

프로그램에 해당된다. 축구 게임 프로그램은 행동 제어부와 행동 전략 생성부의 두 부분으로 구성된다. 사람에 의해 미리 구성되는 행동 제어 부는 축구 로봇의 모터 속도 제어 함수, 위치 이동 함수, 그리고 몇 가지 킥 동작을 수행하는 함수들로 구성된다. 또 ZCS에 의해 동적으로 생성되어지는 행동 전략 구성 부는 언제 공격과 방어를 수행 할 지, 어떤 로봇이 어떤 종류의 킥을 하고 어떤 로봇이 보조를 할지, 또 보조를 하는 로봇은 어떤 위치로 이동 할 것인지를 결정하게 되며 골을 획득할 수 있는 규칙들을 발견함으로써 전략을 생성하게 된다.

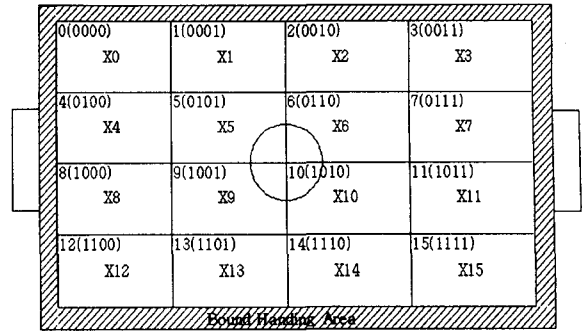


그림 2. 검지기과 효과기를 위한 경기장 지도

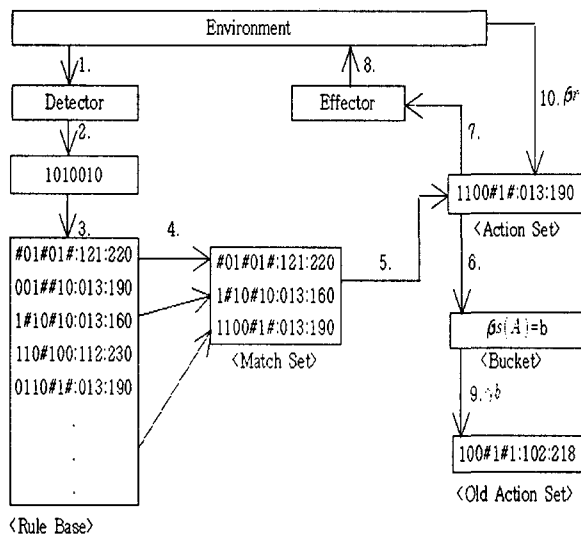


그림 1. 시뮬레이션에 적용된 ZCS의 구조

ZCS으로 구성한 행동 전략 생성 부는 다음과 같다[4].

1. 검지기는 외부 환경의 상태를 7비트의 이진 문자열로 디코딩한다. 디코딩된 메시지의 첫번째 비트는 공으로부터 어느 팀 로봇이 더 가까이 있는지를 나타내는데, 공과 각 로봇과의 거리가 가장 가까운 로봇이 우리 팀 로봇일 경우 0, 상대 팀 로봇일 경우 1로 표시한다. 두 번째 비트는 공이 그림2에 나타난 경기장의 '가 장자리 영역(Bound Handling Area)'에 위치 한 경우 0, 그렇지 않은 경우는 1로 표시 한다. 또 세 번째 비트부터 여섯 번째 비트까지의 4 비트는 그림 2에 나타난 경기장에서 공의 위치를 표시한다. 그리고 마지막 일곱 번째 비트는 공의 이동 방향을 표시하는데 만약 공이 우리 팀 골로부터 상대 팀 골을 향해 이동할 경우 0, 반대의 경우 1로 표시한다.

2. 초기 규칙 베이스의 각 분류자들은 임의로 생성되어지는데, 조건부에 해당되는 7비트는 0과 1 그리고 '#(wild card)'로 구성된다. 다음 네 비트는 행동부를 나타내며 여러 가지 행동 제어 명령에 해당되는 정수로 구성되어진다. 행동부의 첫 번째 비트는 0과 1의 값을 가질 수 있고, 두 번째 비트는 0부터 2까지, 세 번째 비트는 0에서 26까지, 네 번째 비트는 0에서 239까지의 행동 제어 명령을 표현 할 수 있다. 분류자의 마지막 비트에 할당된 값들은 각 분류자들의 신뢰도를 나타내며 초기에 모두 200으로 동일하게 주어진다.

3. 외부 환경으로부터 메시지를 받게 되면 규칙 베이스의 분류자들의 조건부와 비교하여 메시지와 일치하는 조건부를 가진 분류자들로 매치 집합을 만든다. 그리고 이들 매치 집합의 분류자들의 신뢰도를 기준으로 룰렛 선택하여 행동 집합의 분류자를 구성하게 되는데 이 때 매치 집합에서 선택된 행동 집합의 분류자의 행동부와 같은 행동부를 가진 매치 집합의 분류자들을 행동 집합에 추가한다. 또 만약 외부 환경으로부터 받은 메시지와 일치하는 조건부를 가진 분류자가 규칙 베이스에 존재하지 않을 경우는 "covering operation"을 수행하게 되는데 이 때에는 임의로 새로운 분류자를 하나 생성하고 조건부는 외부 상태 메시지의 임의의 위치를 선택하여 "#"로 대체하여 구성한다. 또 행동부는 초기 생성할 때와 마찬가지로 임의로 생성하며 신뢰도는 규칙 베이스의 모든 분류자들의 신뢰도 값의 평균값을 사용한다. 이렇게 생성된 새로운 분류자를 규칙 베이스에 추가하게 되는데 규칙 베이스의 각 분류자들의 신뢰도의 역수를 이용하여 룰렛 선택된 분류자와 대체함으로써 새로운 분류자를 규칙 베이스에 추가하고 다시 매치 집합과 행동

집합을 구성한다.(그림 1의 1에서 5까지)

4. 행동부의 첫번째 비트는 공격 상황인지 수비 상황인지에 따른 행동 명령을 나타낸다. 두 번째 비트는 각 상황에 따라 몇 개의 로봇이 킥을 하고 몇 개의 로봇이 보조를 할 것인지를 나타낸다. 세 번째 비트는 두 번째 비트로부터 역할 수행을 위해 선택된 로봇이 어떤 종류의 킥을 수행 할 것인지를 나타낸다.
5. 행동 집합의 모든 분류자들은 이전의 행동 집합의 분류자들에게 비용을 지불하게 되는데 이 때 사용되는 파라미터들을 정리하여 표 1에 나타내었다. 행동 집합의 모든 분류자들의 개수를 $|A|$, 각 분류자들의 신뢰도 값을 $s(A)$, 이전의 행동 집합의 모든 분류자들의 개수를 $|O|$, 행동 후 외부 환경으로부터 평가된 값을 r 라 하면 $\beta r / |A|$ 이 행동 집합의 분류자들에게 보상 되어지며 내재적 버킷 브리게이드 알고리즘에 의해 $\beta s(A) / |O|$ 이 이전의 행동 집합의 분류자들에게 지불 되어진다.(그림 1의 6에서 10까지)

표 1. ZCS의 각종 파라미터

기호	기호의 설명
β	내재적 버킷 브리게이드 알고리즘에서 신뢰도를 갱신하기 위해 사용되는 Learning rate
γ	이전에 활성화된 행동 집합의 분류자들에게 지불할 비용을 결정할 때 사용되는 Discounter factor
ρ	유전자 알고리즘 적용 확률

6. 위의 과정을 마친 후 현재 평가된 행동 집합은 과거 행동 집합(old action set)에 기록되고 위의 과정을 반복한다. 이 때 확률 ρ 에 의해 유전자 알고리즘이 적용된다. 만약 유전자 알고리즘이 적용된다면 규칙 베이스에 있는 분류자들의 신뢰도 값에 의한 룰렛 선택으로 두개의 부모 분류자가 선택되고 교차와 돌연변이를 거친 후 신뢰도의 역수 값에 의해 룰렛 선택된 두개의 분류자와 대체됨으로써 높은 신뢰도를 가진 분류자들이 규칙 베이스에서 증가하게 된다. 교차는 조건부와 행동부를 경계로 하여 1점 교차를 하였으며 교차 후 생성된 자식 분류자의 신뢰도는 부모 분류자들의 신뢰도의 평균값이 된다.

2.2 ZCS을 이용한 로봇 축구 시뮬레이션

ZCS을 로봇 축구라는 외부 환경에 적용하기 위해서는 “이벤트와 샘플링 시간(event and sampling time)”이라는 개념의 도입이 필요하다. 이것은 일정한 시간 간격동안 행동 집합을 환경 속에서 평가를 하고 보상 값을 획득하는 경우가 발생했을 때만 규칙 베이스의 갱신이 이루어지도록 한다는 것이다. 그림 3에 이벤트의 종류와 이벤트가 발생 했을 때 획득되는 점수를 나타내었다.

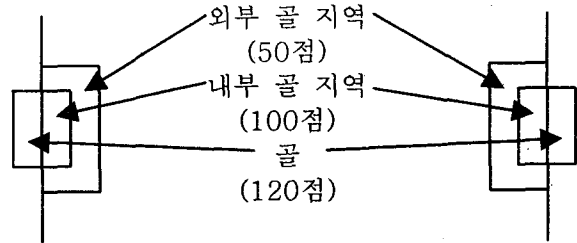


그림 3. 이벤트의 종류와 획득 점수

경기 중 공이 외부로부터 각 영역에 들어가게 될 경우 할당된 만큼의 점수를 얻게 된다. 만약 공이 상대팀 외부 골 지역까지 들어 갔다가 상대팀 로봇에 의해 차단 될 경우 50점을 획득하게 되고 골이 획득되었을 경우 최고 점수인 170점을 얻게 된다. 이와 반대로 우리 팀 영역에 들어오게 되면 해당되는 만큼의 점수를 잃게 된다. 샘플링 시간동안 획득된 점수에 의한 보상 값 $r = \beta \times (120 \times G + 100 \times I + 50 \times O)$ 이 되고 여기서 G 는 샘플링 시간동안 획득한 골의 수를, I 는 샘플링 시간동안 공이 내부 골 지역에 들어가는 이벤트가 발생한 횟수를, O 는 외부 골 지역에 들어가는 이벤트가 발생한 횟수를 각각 나타낸다. 게임을 통한 시뮬레이션은 동일한 행동 제어 함수를 가지지만 사람에 의해 구성된 전략을 보유한 팀과 ZCS으로 전략을 구성하는 팀, 두 팀으로 나누어 진행된다.

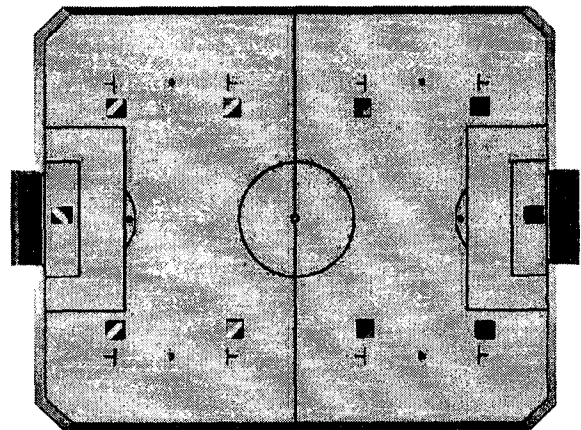


그림 4. 로봇 축구 시뮬레이터

게임 중 적용된 유일한 규칙은 골이 발생했을 경우 그림 4와 같은 위치에 각 로봇과 공을 위치시킨 후 게임을 계속 진행한다. 본 논문에서 사용된 ZCS 파라미터들은 다음과 같다.

$\lambda = 50$, $\beta = 0.2$, $\gamma = 0.05$, $\rho = 1.0$, 교차 확률=0.8, 돌연변이 확률=0.08, 샘플링 시간=3초로 설정하였다.

2.3 시뮬레이션의 결과

그림 5와 6은 594번의 이벤트가 발생할 때까지 규칙 베이스를 구성하는 분류자들의 평균 신뢰도 값의 변화와 보상 값의 변화를 나타낸 그래프이다. 보상 값의 변화를 나타낸 그래프를 보면 원으로 표시된 약 430번째 이벤트가 발생한 이후 구간에서 같은 기간을 가진 다른 구간에 비해 점수를 잃는 횟수가 많이 감소한 것을 알 수 있다. 또 그림 5의 평균 신뢰도 값의 변화를 보면 초기 신뢰도 값에서 점점 큰 폭으로 변화를 하며 전체적인 평균 값이 올라가다가 원으로 표시된 그림 6과 같은 구간에서 300에서 350사이의 값을 유지하면서 다른 구간에 비해 안정되어 가는 것을 확인할 수 있다. 두 그래프를 전체적으로 살펴보면 새로운 규칙을 발견하기 위한 "exploration"과 획득된 유용한 규칙들을 적극적으로 활용하는 "exploit"를 나타내는 구간이 존재함을 알 수 있고 이 과정을 거치면서 외부 환경에 적응하고 좀더 많은 점수를 얻기 위한 유용한 규칙들을 학습해 가는 것을 알 수 있다. 그러나 어느 정도 수렴된 상태에서도 여전히 큰 점수를 잃게 되는 이벤트가 종종 발생하는데 이것은 빠르게 변하는 외부 환경의 상태 변화로 인해 로봇끼리의 충돌이나 이동 중 공과의 충돌과 같은 돌발적인 상황이 발생함으로써 생기는 결과로써 이런 문제를 해결하기 위해서는 각 로봇간의 거리와 같은 좀더 많은 외부 환경의 정보를 규칙 베이스로 전달 해주어야 함을 의미한다.

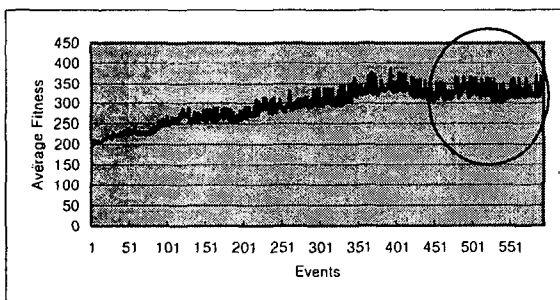


그림 5. 평균 신뢰도(적합도) 변화

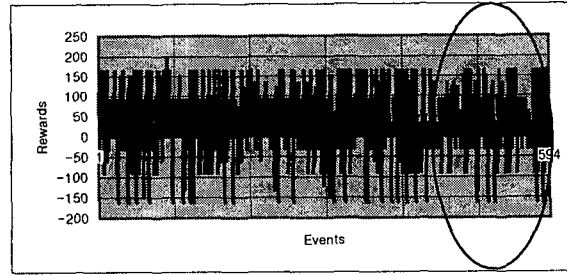


그림 6. 보상 값의 변화

III. 결 론

본 논문에서는 로봇 축구 시뮬레이션 게임 (SimuroSot)의 전략을 ZCS을 이용하여 구성하였으며 로봇 축구 시뮬레이션 프로그램 (RobotSoccer1.2)을 통하여 시뮬레이션 결과를 분석함으로써 ZCS이 상태가 시간에 따라 동적으로 변하는 외부 환경에 적응하고 유용한 행동 규칙을 발견하여 더 나은 전략을 구성할 수 있음을 확인 하였고 이로써 ZCS을 이용하여 다 개체로 구성된 축구 로봇의 전략 생성 시스템을 구현하는 것이 가능함을 알게 되었다.

IV. 참고문헌

- [1] David E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, New York, Addison-Wesley, 1989.
- [2] Jacob Hurst, Larry Bull and Chris Melhuish, "ZCS and TCS Learning Classifier System Controllers on Real Robots," *PUWE Learning Classifier System Group Technical Report - UWELCSG02-002*, 2002.
- [3] Larry Bull & Jacob Hurst, "ZCS : Theory and Practice," *PUWE Learning Classifier System Group Technical Report - UWELCSG01-001*, 2002.
- [4] Gary William Flake, *The Computational Beauty of Nature*, Massachusetts, A Bradford Book The MIT Press, 1995.