**공격형 에이전트 보상 함수 설계**

공격형 에이전트의 목표는 적극적으로 상대를 공격하여 승리하는 것입니다. 불필요한 수비 행동을 최소화하고, 효율적으로 적을 제압하도록 유도

* **긍정적(+) 보상:**
  + **적 공격 성공 시:** +0.1
    - *설계 의도:* 에이전트가 공격적인 행동을 하도록 유도하는 가장 중요하고 빈번한 보상
  + **경기 승리 시 (상대 체력 0):** +1.0
    - *설계 의도:* 수비와 반격을 통해 결국 승리하는 것이 최종 목표임을 학습
  + **반격(받아치기) 성공 시:** +0.2
    - *설계 의도:* 위험을 감수하고 방어 후 공격으로 전환하는 효율적인 행동을 장려
* **부정적(-) 보상:**
  + **불필요한 방어/회피 시:** -0.05
    - *설계 의도:* 적의 공격 의도가 없는데도 방어나 회피를 사용하는 소극적인 행동에 페널티를 부과, 자원을 낭비하지 않도록 학습
  + **피해를 입었을 때:** -0.1
    - *설계 의도:* 자신의 체력도 관리해야 함을 인지, 무모한 공격을 피하도록
  + **경기 패배 시 (자신 체력 0):** -1.0
    - *설계 의도:* 패배를 피하도록 강력하게 학습
  + **시간 경과 시 (매 스텝마다):** -0.0001
    - *설계 의도:* 에이전트가 망설이거나 아무것도 하지 않는 행동을 방지, 적극적으로 상황을 해결하도록 유도

**수비형 에이전트 보상 함수 설계**

수비형 에이전트의 목표는 적의 공격을 효과적으로 막거나 피하고, 안전하게 생존하며 반격의 기회를 노리는 것

* **긍정적(+) 보상:**
  + **공격 방어/회피 성공 시:** +0.2
    - *설계 의도:* 수비형 에이전트의 핵심 역할인 방어와 회피 행동을 성공했을 때 즉각적인 보상을 주어, 방어 기술 사용을 장려
  + **효과적인 반격 (받아치기) 성공 시:** +0.3
    - *설계 의도:* 단순히 막는 것에서 그치지 않고, 방어 후 공격으로 전환하는 고효율 행동에 큰 보상을 주어 학습을 유도
  + **체력 유지에 비례한 보상:** 에피소드 종료 시 현재 체력 / 최대 체력 \* 0.5
    - *설계 의도:* 게임에서 승리하지 못하더라도, 체력을 잘 보존한 채 에피소드가 끝났다면 긍정적인 보상을 주어 생존의 가치를 학습
  + **경기 승리 시 (상대 체력 0):** +1.0
    - *설계 의도:* 수비와 반격을 통해 결국 승리하는 것이 최종 목표임을 학습
* **부정적(-) 보상:**
  + **피해를 입었을 때:** -0.2
    - *설계 의도:* 공격형 에이전트보다 더 큰 페널티를 부과, 피해를 입는 것을 최대한 피하도록 학습
  + **방어/회피 실패 시 (스킬을 사용했으나 맞았을 경우):** -0.1
    - *설계 의도:* 섣부른 방어나 회피 남용을 방지, 정확한 타이밍에 기술을 사용하도록 유도
  + **경기 패배 시 (자신 체력 0):** -1.0
    - *설계 의도:* 패배를 피하도록 강력하게 학습
  + **시간 경과 시 (매 스텝마다):** -0.0001
    - *설계 의도:* 소극적으로 방어만 하며 시간을 끄는 행동을 방지