# 강화학습 기반 실시간 반응형 퀘스트 생성 시스템 중앙 관리자 영향력 연구

김태훈 [1], 김창재 [2]
[1] 선문대학교 AI 소프트웨어과 학부생
[2] 선문대학교 컴퓨터공학부 교수

raflereak@gmail.com, winchang@sunmoon.ac.kr

# Proposal Realtime Reaction Generate Quest System Basement Reinforcement Learning Central System

Kim-Tae Hun[1], Kim-Chang Jae[2]
[1]Dept. of AI Software, Sun-Moon University
[2]Dept. of Computer Engineering, Sun-Moon University

요 약

강화학습 기반의 다중 에이전트 시스템을 이용한 서버의 실시간 상황을 제공 받아서 상황에 알맞은 퀘스트를 생성해주는 시스템을 제안한다. 학습 가이드 역할을 하는 CTDE 의 중앙 관리자의 역할을 위한 에이전트를 분리하여 작동하게 함으로서 퀘스트의 지향점을 잡는 것이다.

## 1. 서론

본 연구에서는 게임 시스템에 강화학습 기반의 인공지능이 접목되어 수작업으로 진행되던 퀘스트 생성을 대체할 방법을 제안하고자 한다. DQN 강화학습 알고리즘을 다중 에이전트로 변환하여 Multi-Agent 기반의 CTDE 기법을 채택한 강화학습 알고리즘을 이용하며 행동에 대한 피드백이 사용자에 의하여 나중에 반영되기 때문에 PK/PD 기법을 활용한다. 강화학습은 가상환경의 에이전트의 행동에 대해서 보상을 책정하여 학습시키는 방식으로 보상을 통해서 정책을 개선하는 지능형 기계학습 방법이다. 기존 DQN 은 강화학습의 최초 모델로서 다른 강화학습 알고리즘 보다 성능이 뒤쳐진다. 이를 이용하여 Multi-Agent 기법으로도 여러가지가 있는데, 그중 많은 곳에서 채택중인 CTDE 기법을 채택하여 각 에이전트들이 한 방향으로 학습이 될 수 있도록 유도하는 방법을 제안하고 있다. CTDE(Centralized Training with Decentralized Execution)는 전체 정보를 볼 수 있는 중앙 관리자가 행동의 가치를 판단하여 Agent 의 행동 정책을 평가한다. 이에 따라 학습을 진행한 Agent 는 중앙 관리자의 지향성에 맞춰진다. 또한 여러 Agent 들의 기여도를 추정하여 학습의 필요성을 제공하여 정책의 문제점을 인지하게 만든다.

본 연구에서는 각 게임 서버에서 제공한 서버 상태에 대해서 Agent 들이 어떤 정책을 책정할지 기준을 가르는 중앙 관리자의 평가 기준을 세우는 솔루션을 제안한다.

게임을 플레이하는 플레이어에게 있어서 몰입도는 게임을 이용하게 만드는 요인이다. 게임은 많은 발전이 있었고, 플레이어 또한 많은 발전이 있었다. 플레이어의 게임 시스템에 대한 지능 또한 과거에 비해서 비약적으로 상승하였다. 게임의 시스템을 악용하는 사례들이 발생하고 있다.

## 2. 스폰 시스템의 문제점

악용 사례 중 스폰 시스템이 있다. 스폰은 어떤 오브젝트를 컨텐츠 제공의 목적을 위하여 재 생성시켜주는 것으로 모든 게임에서 오브젝트를 다루기 위하여 채택하고 있는 시스템이다. 시스템의 목적은 플레이어들의 컨텐츠 제공의 목적이다. 그러나, 이 시스템은 문제점을 갖고 있다. 이는 서버 및 클라이언트의 부하를 막기 위해서 스폰할 수 있는 최대 오브젝트 수를 제한한다.

이 시스템은 많은 게임을 이용한 플레이어들에게는 몰입도를 떨구는 시스템이며, 악용할 수 있게 만든다. 몹들의 경우 상대하기 쉬운 몹들을 스폰시켜 상대하기 어려운 몹들이 스폰되지 않게 만든다. 이는 오브젝트의 수가 제한되어 스폰되지 않게 되는 것을 이용한 것이다. 허점을 이용하는 사례가 발생하기 시작했다.

대응하기 위해서 관리자를 둘 수 없다. 유기적으로 반응하기에는 너무나도 많은 변수들이 존재한다는 것인데, 수많은 오브젝트의 동향을 파악해서 스폰의 정도를 조절한다는 것은 수작업으로 불가능한 것이다. 이를 해결하기 위한 시스템으로 강화학습 기반 반응

형 퀘스트 생성 시스템을 제안하고자 한다.

## 3. 연구/개발 환경

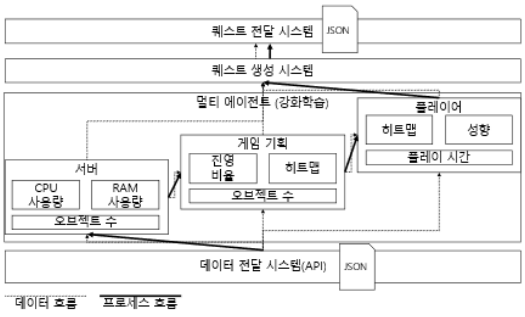
기존 시스템의 문제점이었던 최대 스폰 횟수의 제약을 없애 플레이어들의 몰입도 증강 및 플레이어들의 허점 이용 사례를 막아낸다. 이렇게 된다면 서버와 클라이언트의 부하가 발생하는데, 발생하기 전에 부하를 방지하기 위해서 스폰된 오브젝트들을 제거하면 된다. 이 또한 플레이어들에게 맡기는 것으로 해결하는 것이 시스템의 주된 목적이다. 플레이어들은 게임을 즐기기 위함으로 플레이한다. 이는 단순이 제거하는 것만으로는 즐기기 위한 목적에 충족시킬 수 없다. 따라서 게임은 퀘스트 시스템을 도입하여, 플레이어들에게 목적성 부여 및 달성감에 대한 성취감을 제공하여 플레이를 유도하게 만들어졌다. 이 시스템의 목적대로 퀘스트를 생성하여 플레이어들의 즐기기 위한 목적을 충족시키고, 서버 및 클라이언트의 부하를 미연에 방지할 수 있게 만드는 것이다.

연구를 위해 제안하는 환경은 가상 시뮬레이션을 위한 서버 시뮬레이션 시스템으로 내부 환경은 멀티플레이 목적으로 개발되어 데디케이티드 서버 환경을 채택하였다. 이는 플레이어가 아닌 운영 쪽에서 서버를 운용하게 만들어진 시스템이다. 시스템에는 레벨이라는 가상의 환경이 존재한다. 플레이어들은 맵이라고 부르는 레벨 시스템은 여타 오픈월드 기반의 게임들과 유사한 디자인을 채택하고 있다. 이 위에 스폰 매니저가 있어, 레벨 내에 NPC, Mob, Item 이 지속적으로 생성되도록 만들어준다. 플레이어 AI 를 일정량 생성하여 API 통신 기능이 탑재 된 NPC 와 상호작용하게 만들었다. 오픈월드 기반의 MMO 게임은 많은 플레이어가 동시다발적으로 플레이하는 것이 기본적인 환경이지만 연구개발을 목적으로 개발된 게임 시스템에 많은 플레이어를 유치할 수 없기 때문에 대신 플레이어를 대체할 인공지능을 구현하였다. 본 연구에서는 퀘스트 생성 시스템에 접목된 최초의 DQN 에이전트의 행동이 CTDE 관리자의 평가에 도움이 될 수 있는지 파악하기 위함이기 때문에 가상의 환경이 적용되었음을 밝히는 바이다. 오픈월드 게임에서도 퀘스트는 존재하지만 반응형 퀘스트 생성 시스템 적용을 위하여 API 통신을 통한 Json 데이터 기반 퀘스트 생성 시스템을 구현하였다.

## 4. 시스템 구성

총 6 개의 에이전트들을 통해서 하나의 퀘스트가 생성된다.

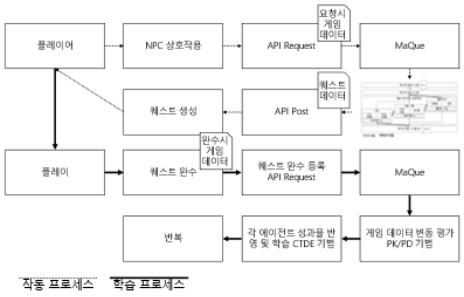
퀘스트 대상, 대상에 대한 대응, 대응 횟수, 대상에 대해서 대응해야하는 위치들을 종합해서 Json 형태로 생성된 퀘스트를 Request 하였던 API 를 통해서 반환해준다. 반환된 Json 데이터는 약속된 형태의 것으로 내용 그대로 퀘스트를 생성하여 요청한 플레이어에게 퀘스트를 만들어 제공한다.




(그림 1) 강화학습 기반 시스템 아키텍처.

아래의 그림은 플레이어가 퀘스트를 요청하면서 발생하는 일반적인 시나리오에 대한 그림이다.

플레이어는 게임 내에 존재하는 NPC 에게 인터렉트를 실행할 수 있다. 이때 NPC 는 본 연구의 시스템에 API 를 통해서 퀘스트 생성을 위한 Json 를 Request 하며 서버 및 그 외로 요구되는 데이터들을 Json 형태로 Post 한다. 요청을 받은 직후 이전의 내용대로 에이전트들이 독립변수들을 이용하여 행동을 진행한다. 그렇게 만들어진 Json 을 반환하여 게임 시스템에 의해서 퀘스트가 생성되고 플레이어가 퀘스트를 수행할 수 있도록 제공된다.




(그림 2) 시스템 프로세스.

플레이어는 퀘스트를 수령하여 시스템에서 제공하는 정보대로 수행한다. 일정 시간이 지난 후에 퀘스트를 완수한다면 다시금 API 를 통해서 이전에 보냈던 데이터들을 변화한 현재의 데이터들을 Json 으로 Post 한다. 이는 변화한 서버의 상태를 피드백으로 받아들여 중앙 관리자 및 각 에이전트들의 행동 정책에 대해서 평가를 진행한다.

## 참고문헌

[1] 유병현, 멀티 에이전트 강화학습 기술 동향, 전자통신동향분석, 35 권, 제 6 호, 138 쪽, 2020 년

[2] 김환희, 신경망은 콘텐츠 자동 생성의 꿈을 꾸는가, NDC, 2016