Paper Title

김신영(18)1, 오상훈(18)2

*1*Author(s) Affiliation(s)

*2Author(s) Affiliation(s)*

email@xx.xx1*, dec@xx.xx2*

# *Abstract*

The abstract should summarize the contents of the paper and written below the author information. Use the word “Abstract” as the title, in 12-point Times New Roman, boldface type, italicized, centered relative to the column, initially capitalized, fixed-spacing at 11 pt. spacing before the text and 0 pt. after. The abstract content is to be in 11- point, italicized, single spaced type. Leave one blank line after the abstract, and then begin the keywords. All manuscripts must be in English.

***Keywords:*** *Write more than 4-6 Keywords.*

1. Introduction

Advanced Engineering and ICT-Convergence Proceedings (AEICP)

ISSN : 2635-4586

©ICTAES 2018

컴퓨터 게임은 강화학습 연구 분야에서 높은 관심을 받는 환경 중 하나이다. 이러한 게임은 다양한 요소와 도전 과제를 제공하며 현실 세계의 다양한 응용 분야와 유사한 문제를 모델링한다. 특히 스타크래프트2는 실시간 전략 게임으로, 강화학습 연구의 중요한 벤치마크로 인정받고 있다. 이 게임은 광범위하고 다양한 행동 공간과 다중 에이전트 환경을 제공하며 불완전한 정보를 가지고 있어, 강화학습 알고리즘의 일반화 능력을 테스트하고 불확실하고 다양한 정보 아래에서 최적의 결정을 내려야 하는 상황을 모델링할 수 있다.

기존 스타크래프트2 관련 연구들은 강화학습을 통해 게임의 마스터리를 시도하거나 다중 에이전트 강화학습을 활용하여 그랜드 마스터 수준의 성과를 달성한 연구가 있다. 이러한 연구들은 스타크래프트2를 강화학습 연구의 중요한 대상으로 인정하고 있으며, 게임 인공지능 분야에서 큰 관심을 받고 있다.

본 연구에서는 스타크래프트 2 환경에서 강화학습 알고리즘을 활용하여 최적의 빌드 전략을 개발하고 이를 통해 특정 유닛을 가장 효율적으로 생산하여 전투에서 승리하는 방법을 탐구할 것이다. 이 빌드 전략은 게임에서 승리에 중요한 역할을 하며, 각 빌드에는 다양한 전략적 선택과 게임 상황에 대한 대응이 포함되어 있다. 이를 통해 스타크래프트 2를 통한 강화학습 연구의 중요성과 가능성을 강조하고, 게임 전략 개발 및 게임 인공지능 분야에 새로운 기여를 하고자 한다.

2. Background

스타크래프트 2는 강화학습 연구 분야에서 혁신적인 환경을 제공하며, 컴퓨터 게임이 강화학습 연구에 미치는 중요성을 강조한다. 이 게임은 실시간 전략 환경을 가지며, 다양한 종족, 유닛, 건물, 리소스 등이 상호작용하고 복잡한 규칙이 적용되는 특징을 가진다. 스타크래프트 2는 강화학습 연구에 이상적인 환경을 제공하는데, 이를 통해 강화학습 알고리즘의 다양한 측면을 검증하고 현실 세계 응용 분야와 유사한 문제를 모델링할 수 있다.

먼저, 스타크래프트 2는 다양한 가능한 행동을 제공하며, 이로 인해 강화학습 알고리즘은 다양한 상황에서 최적의 행동을 선택해야 한다. 게임 내에는 여러 플레이어나 AI 유닛이 경쟁하거나 협력하는 다중 에이전트 환경이 구성되어 있어, 이러한 다중 에이전트 상호작용에서 강화학습은 다양한 전략을 고려하고 최적의 결정을 내리는 데 활용된다. 뿐만 아니라 플레이어는 게임 내 정보의 부분적인 공개로 인해 불완전한 정보 상에서 의사 결정을 내려야 한다.

스타크래프트 2를 통한 연구는 게임 인공지능 분야에 혁신을 가져오며, 강화학습 알고리즘의 일반화 능력을 검증하고 다양한 응용 분야에 적용할 수 있는 중요한 연구 분야로 간주된다.[1] 더 나아가, 스타크래프트 2에서 강화학습을 적용하여 게임 전략을 개발하고 최적화하는 연구는 게임 인공지능 분야에 새로운 기여를 하며, 이러한 연구가 강화학습의 가능성을 강조한다.

이어서, 스타크래프트 2를 활용한 연구 중에서는 AI 에이전트가 게임에서 우수한 성과를 달성하는 혁신적인 AlphaStar가 나왔다.[2] AlphaStar는 강화학습과 신경망 기술을 활용하여 게임 전략을 학습하고 경쟁 상대를 이기는 방법을 개발한 프로젝트다. 이 프로젝트는 높은 차원의 행동 공간과 다중 에이전트 환경에서 작동하는 AI를 개발하는 데 있어서 중요한 성과를 이루었다.

이후에는 스타크래프트 2에서 다중 에이전트 강화학습을 활용하여 그랜드 마스터 수준의 성과를 달성하는 방법 또한 나왔다.[3] 스타크래프트 2의 다중 에이전트 환경이 강화학습 연구에 적합하다는 점을 강조하며, 다중 에이전트 강화학습 알고리즘을 통해 경기력을 향상시키는 연구가 나온 것이다.

스타크래프트 2를 강화학습 연구에 활용한 많은 선행 연구가 있으며, 본 연구는 이러한 연구를 바탕으로 스타크래프트 2에서의 최적화된 빌드 전략과 최단 시간 내에 승리하는 전투 전략을 개발하는 방법에 중점을 두고 있다.

**3.** Methodology

본 연구에서는 스타크래프트2 내 빌드 전략과 전투 전략을 개발했다. 멀티가 하나인 가정하에서 특정 시점에 가장 많은 미네랄과 가스를 획득하도록 하는 것을 강화 학습을 통해 계산하였다. 강화 학습의 평가는 동일한 맵과 초기 상태에서 시작하며 마르코프 연쇄 법칙을 활용하여 이론적으로 계산한 값과 강화 학습 결과를 비교하였다.

먼저, 미네랄 수급률을 최대화하는 빌드 전략을 강화학습을 통해 계산하였다. 미네랄 수급률에 대한 공식은 미네랄 단위당 일꾼 수가 2배수일 때 최적 효율이라는 것을 고려되었다. 이를 바탕으로 초기 빌드의 상성을 계산하고, 유닛 생산 시간 및 이동 거리와 같은 다른 변수도 고려하여 초기 빌드를 최적화하고 플레이어의 피로감을 줄일 수 있는 전략을 개발하였다.

미네랄 수급 봇을 개발하는 목적은 복잡한 실시간 전략 게임인 스타크래프트2의 게임 환경을 이해하고 에이전트가 현재 상태(미네랄 양, 유닛 수 등)를 고려하여 최적의 미네랄 수집 행동을 학습하도록 하는 것이었다. 이를 위해 봇 학습을 위한 알고리즘으로 Q-Learning 알고리즘을 사용하였다.[5]

**3-1. 비동기 함수**

**3-2. Q-Learning**

Q Learning은 강화학습 알고리즘 중 하나로, 상태(state)와 행동(action)의 조합에 따른 보상의 기대치를 나타내는 Q-값을 학습하는데 초점을 맞춘다. 이 알고리즘은 각 상태와 행동에 대한 최적의 Q-값을 찾아내기 위해 사용됩니다. 다음은 Q-Learning에 대한 주요 내용을 간략하게 정리한 것입니다:

기반이기 때문에 이전에 논의한 Lookup-Table 기반 접근 방식 중

강화 학습의 가치 기반 방법론 중 하나로, 특정 상태에서의 행동의 가치를 추정하는 Q-value를 중심으로 한다. 이 Q-value는 Q(s,a)로 나타낼 수 있다. 여기서 s는 상태, a는 행동을 나타낸다. Q-learning의 기본 모델에서는 상태와 행동 간의 모든 조합에 대한 Q-value를 저장하는 Q-Table이라는 데이터 구조를 사용한다. Q-Learning의 핵심 공식은 Bellman Equation을 기반으로 하는 핵심 공식을 통해 업데이트된다. Q-value 핵심 공식은 다음과 같다. Q(s,a)←(1−α)×Q(s,a)+α×(r+γ×maxa′​Q(s′,a′))이다. 여기서 α는 learning rate로 새로운 Q-value 추정치를 이전값에 얼마나 반영할 것인지를 결정하고 γ는 discount factor로 미래의 보상을 현재 보상에 얼마나 중요하게 여길 것인지를 나타낸다. 또한 r는 현재 상태와 행동에 대한 보상, s′는 다음 상태를 나타내고, max′​Q(s′,a′)는 다음 상태 s′에서 가능한 모든 행동 a′에 대한 최대 Q-value를 나타낸다.

Q-Learning에서는 Exploration과 Exploitation의 균형을 유지하는 것이 중요하다. 이를 위해 ϵ-greedy 전략을 사용한다. 확률 ϵ로 무작위 행동을 선택 (Exploration)하고, 확률1−ϵ로 현재 Q-value가 가장 높은 행동을 선택 (Exploitation)한다. 시간이 지남에 따라 ϵ은 줄어들어서 점점 더 많이 활용하는 전략으로 전환될 수 있다.

Q-Learning은 환경 모델을 필요로 하지 않고 게임 내 상황에 따라 병렬처리를 통해 여러 상태와 행동 조합을 동시에 학습할 수 있는 강화 학습 기법 중 하나이다. 따라서, 스타크래프트2와 같이 복잡한 게임에 적용하기에 적합하다 판단했다. 스타크래프트2에서의 reward function은 다음과 같이 설정했다. 미네랄 수집 최적화의 목적으로 강화 학습을 진행한다면:

+1의 보상, -5의 보상

등으로 설정할 수 있습니다. 그리고 이 reward function을 통해 에이전트가 특정 상태에서 어떤 행동을 선택해야 하는지를 학습하는데 사용되었다.

Q-Learning에서는 일반적으로 Mean Squared Error(MSE)를 손실 함수로 사용하여 Q-value의 예측값과 실제값 사이의 차이를 최소화하려고 한다. MSE=1/N​∑(Q(s,a)predicted​−Q(s,a)true​)^2. 여기서 N은 배치 크기를 나타낸다. 이러한 설정을 통해 에이전트는 각 상태와 행동에서 예상되는 미래 보상을 최대화하도록 학습되었다. 이 과정은 일정한 성능 향상이 이루어지거나, 최대 반복 횟수에 도달할 때까지 반복적으로 진행된다

스타크래프트2 게임에서는 다양한 환경 요소를 고려하여 게임 환경을 모델링하고 데이터를 수집하여 학습했다. 스타크래프트2 게임 환경 모델링에는 유닛의 움직임, 미네랄 위치, 적군 유닛 및 구조물 위치 등 다양한 정보가 사용되었다. 또한, 시간과 관련된 정보를 확인하고 보상으로 활용하기 위해 self.check 변수를 통해 사용했다.

Q-Learning 알고리즘의 성능 최적화를 위해 학습률, 할인 요소, 엡실론 값 등 하이퍼파라미터를 조정하였다. 성능은 목표 미네랄량 달성 여부를 기준으로 평가되었으며 게임 진행 시간, 미네랄 채집 속도, 승리 확률 등 다양한 지표를 통해 분석되었다. 이를 통해 Q-Learning 기반의 미네랄 수집 봇의 효과가 확인되었다.

다음으로는 'IncrediBot'의 핵심 로직을 포함하는 봇이 프로토스 종족으로 게임을 진행하며, 게임 상황을 실시간으로 모니터링하고, 적절한 전략을 선택하여 최단 시간에 승리하는 전투 전략을 강화학습을 통해 계산하였다.

초기 설정으로 게임 맵은 "Simple64"를 사용하였고, 상대방은 마찬가지로 프로토스 종족의 AI인 "Computer"로 설정하였다. 봇은 초기 단계에서 게임 시작 시 자원 및 유닛 상태를 확인하고 초기 설정을 수행한다. 이를 통해 자원 생산과 유닛 훈련에 필요한 정보를 파악한다.

본 논문에서는 봇 구현을 위해 python-sc2 라이브러리를 사용하였다.[4] 이 라이브러리는 StarCraft II의 AI 개발을 돕기 위해 제공되며, Python 환경에서 강화 학습 등의 기법을 통해 게임 내 에이전트를 손쉽게 구현할 수 있게 해준다.

Gateway, Cybernetics Core, Stargate

스타크래프트2 환경에서 강화 학습을 적용하려면 다양한 환경 요소를 고려해야 합니다. 예를 들면, 유닛의 움직임, 미네랄 위치 등이 있습니다. 본 연구에서는 Void Ray 유닛 5대가 생성되는 시간을 중심으로 보상 시스템을 설계하였습니다. 이를 위해 iteration을 활용하여 보상을 정의하였고, 보상은 에이전트가 Void Ray를 빠르게 생산하는 데 얼마나 성공적인지에 따라 부여되었습니다.

일반적인 상황에서, 사람이 시도할 때 정해진 빌드를 따를 경우 Void Ray 5대를 생산하는 데 약 6분이 걸릴 것이라는 가정하에 시작하였습니다. 그러나 일꾼을 추가로 생산하지 않고 단 하나의 Stargate만을 건설한다는 제한된 환경에서 Q-Learning을 적용하였을 때, 5분 36초 만에 Void Ray 5대를 생산하는 결과를 얻을 수 있었습니다. 이는 기존 가정보다 빠른 시간입니다. 이러한 성능 향상을 위해 사용자가 임의로 제한 조건을 설정하는 프루닝 기법을 사용하여 시간을 최적화했다. 또한 여러개의 Stargate, 여러 개의 일꾼을 설정한 경우 멀티의 경우도 고려해야 한다는 등 사람이 예측하기 어렸웠던 경우도 Q-Learning을 통해 알 수 있었습니다.

결론적으로, Q-Learning 방법론은 복잡한 게임 환경인 스타크래프트2에서도 효과적으로 적용될 수 있음을 확인하였습니다. 특히, 주어진 제한 조건과 보상 시스템 하에서 에이전트가 최적의 전략을 성공적으로 학습하였으며, 이를 통해 게임 내에서의 성능을 효과적으로 향상시킬 수 있음을 보여주었습니다.

**4. Conclusion**

본 연구에서는 스타크래프트2를 강화학습 연구의 중요한 대상으로 제시하고 게임 내 빌드 전략 및 전투 전략을 개발하여 그 가치를 입증하였다. 스타크래프트2 환경은 강화학습 알고리즘의 일반화 능력을 검증하고, 불확실하고 다양한 정보 아래에서 최적의 결정을 내리는 능력을 모델링하는데 매우 적합한 것으로 나타났다.

이러한 연구를 통해 게임 환경에서의 강화학습 연구의 중요성과 가능성을 강조하며, 게임 인공지능 분야에 새로운 기여를 제공함을 확인했다. 미네랄 수집과 전투 전략 최적화를 통해 게임에서의 인공지능 성능을 향상시킬 수 있음을 입증하였으며, 앞으로의 연구에서는 이러한 결과를 발전시켜 게임 인공지능 분야에 더 큰 기여를 할 것으로 기대한다.

Table 1. Main parameters

|  |  |  |
| --- | --- | --- |
| **Heading level** | **Example** | **Font size and style** |
| Title (centered) | Lecture Notes … | 14 point, bold |
| 1st-level heading | 1 Introduction | 13 point, bold |
| 2nd-level heading | *2.1 Printing Area* | 11 point, italic |
| 3rd-level heading | *Remark.* Text follows … | 11 point, italic |

4.1. Figures

Grids and details within the figures must be clearly legible and may not be written one on top of the other. Figures should be numbered and should have a caption which should always be positioned under the figures, in contrast to the caption belonging to a table, which should always appear above the table. Please center the captions between the margins and set them in 9-point type and Times New Romans font. The distance between figure and caption should be about 6 pt, and the distance between caption and text about 10 pt. Use of 600 dpi/1200 dpi for illustrations for clarity recommended.

Please be noted that illustrations are black and white in printed volumes. However, colored pictures will be shown in the electronic version. If you send colored figures that are to be printed in black and white, please make sure that they really are legible in black and white. Some colors show up very poorly when printed in black and white.



Figure 1. Sample.

*4.2. Equation*

The equation should be written by using MS word as follows:

(1)

*4.3. Units, Equations and Symbols*

For units, the SI system should be used for all dimensional quantities. Leave a space between numbers and units except oC and % (e.g., 40 mg/L, 50%, 10 h).

Main symbols used in equations and text should be in italic. Equations must be numbered consecutively using Arabic numbers.

**5. Conclusion**

Please be noted that illustrations are black and white in printed volumes. However, colored pictures will be shown in the electronic version. If you send colored figures that are to be printed in black and white, please make sure that they really are legible in black and white. Some colors show up very poorly when printed in black and white.

**Acknowledgement**

Acknowledgement title is not numbered. Type the acknowledgement in this format ‘This work was supported by (institution to acknowledge) in (year).’

Note: Manuscripts in which references are not in this format will be returned without review.

**References**

1. Oriol Vinyals, Timo Ewalds, Sergey Bartunov (2017). StarCraft II: A New Challenge for Reinforcement Learning.
2. The AlphaStar team (2019). AlphaStar: Mastering the real-time strategy game StarCraft II.
3. Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature volume 575, 350–354
4. <https://github.com/BurnySc2/python-sc2/tree/develop>
5. 고민수 외 1명, 2023, 『Deep Learning Bible - 5. Reinforcement Learning』, e-book , https://wikidocs.net/165849