Mastering Basketball with Deep Reinforcement Learning: An Integrated Curriculum Training Approach*

Extended Abstract

Hangtian Jia 1, Chunxu Ren 1, Yujing Hu 1, Yingfeng Chen 1,*, Tangjie Lv 1, Changjie Fan 1
Hongyao Tang 2, Jianye Hao 2
1Netease Fuxi AI Lab, 2Tianjin University
{jiahangtian,renchunxu,huyujing,chenyingfeng1,hzlvtangjie,fanchangjie}@corp.netease.com
{bluecontra,jianye.hao}@tju.edu.cn

ABSTRACT

Despite the success of deep reinforcement learning in a variety type of games such as Board games, RTS, FPS, and MOBA games, sports games (SPG) like basketball have been seldom studied. Basketball is one of the most popular and challenging sports games due to its long-time horizon, sparse rewards, complex game rules, and multiple roles with different capabilities. Although these problems could be partially alleviated by common methods like hierarchical reinforcement learning through a decomposition of the whole game into several subtasks based on game rules (such as attack, defense), these methods tend to ignore the strong correlations between these subtasks and could have difficulty in generating reasonable policies across the whole basketball match. Besides, the existence of multiple agents adds extra challenges to such game. In this work, we propose an integrated curriculum training approach (ICTA) which is composed of two parts. The first part is for handling the correlated subtasks from the perspective of a single player, which contains several weighted cascading curriculum learners that can smoothly unify the base curriculum training of corresponding subtasks together using a Q-value backup mechanism with a weight factor. The second part is for enhancing the cooperation ability of the basketball team, which is a curriculum switcher that focuses on learning the switch of the cooperative curriculum within one team by taking over collaborative actions such as passing from a single-player’s action spaces. Our method is then applied to a commercial online basketball game named Fever Basketball (FB). Results show that ICTA significantly outperforms the built-in AI and reaches up to around 70% win-rate than online human players during a 300-day evaluation period.

KEYWORDS

basketball, reinforcement learning, curriculum learning

1 INTRODUCTION AND RELATED WORK

Although different kinds of games have variant properties and styles, deep reinforcement learning (DRL) has conquered a variety of them through different techniques, such as DQN for Atari games [11, 12], Monte Carlo tree search (MCTS) for board games [15, 16], game theory for card games [2], curriculum learning (CL) for first-person shooting (FPS) games [8], hierarchical reinforcement learning (HRL) for multiplayer online battle arena (MOBA) games [5, 13] and real-time strategy (RTS) games [10, 18]. However, sports games (SPG) have rarely been studied except that Google released a football simulation platform. However, this platform only provided several benchmarks with baseline algorithms without proposing any general solutions [7].

![Figure 1: Correlations of Subtasks in Basketball](image)

As another popular sports game, basketball represents one special kind of comprehensive challenge that unifies most of the critical problems for RL. First of all, the long-time horizon and sparse rewards properties remain an issue to modern DRL algorithms. In basketball, it is normally not until scoring shall the agents get a reward signal (goal in or not), which requires a long sequence of consecutive events such as dribbling and breaking through the defense of opponents. Second, a basketball game is composed of many different sub-tasks according to game rules (Figure 1), for example, the offense sub-task (attack, assist), the defense sub-task, the sub-task of acquiring ball possession (ballclear), and the sub-task of navigation to the ball (freeball), each of which can be independently formulated as a RL problem. Third, it is a multi-agent system that needs the players in a team to cooperate well to win the game. The last but not least, there are several characters or positions classified based on the players’ capabilities or tactical strategies such as center (C), power forward (PF), small forward (SF), point guard (PG), shoot guard (SG), which adds extra stochasticity to the game.

For solving complex problems, hierarchical reinforcement learning (HRL) is commonly used by training a higher-level agent to
generate policies on manipulating the transitions of those sub-tasks [1, 6, 9, 14, 17]. However, HRL is not feasible here since these transitions are controlled by the basketball game. Besides, the correlations between sub-tasks (Figure 1) make it more challenging to generate a reasonable policy across the whole basketball match since policies for these sub-tasks are highly correlated.

2 ALGORITHM

By regarding the five subtasks as base curriculums for basketball playing, we propose an integrated curriculum training approach (ICTA) which is composed of two parts. The first part is for handling the correlated subtasks from the perspective of a single player, which contains several weighted cascading curriculum learners of corresponding subtasks. Such training can smoothly establish relationships during the base curriculum training of correlated subtasks (such as $\tau_i$, $\tau_j$) by using a Q-value backup mechanism when calculating the Q-label $y_{\tau_i}$ [11] of the former subtask $\tau_i$ and heuristically adjusting the weight factor $\eta \in [0, 1]$:

$$y_{\tau_i} = \begin{cases} r_i + \eta \max_{a} Q_{\tau_j}^* (s'_{\tau_j}, a'_{\tau_j}, \theta_j), & \text{terminal } s \text{ of } \tau_i \\ r_i + \eta \max_{a} Q^* (s_{\tau_i}, a_{\tau_i}, \theta_i), & \text{otherwise} \end{cases}$$

The second part is a curriculum switcher that targets at enhancing the cooperation ability of the whole team. It focuses on learning the switch of the cooperative curriculum within one team by taking over collaborative actions such as passing from single-player’s action spaces (Figure 2). The switcher has a relatively higher priority over those base curriculum learners on action selection to ensure the performance of coordination within a team.

3 EXPERIMENTAL RESULTS

We evaluate our models in a commercial online basketball game named Fever Basketball (FB). The ablation experiments (Figure 3 (a)) are first carried out by training with the rule-based built-in AI that has average human capability to demonstrate the contributions of our methods. The results of a 300-day online evaluation with 1, 130, 926 human players are illustrated in Figure 3(b). The base algorithm we used is APEX-Rainbow [3, 4].

4 DISCUSSION AND CONCLUSIONS

In the ablation experiments, we can find that the one-model training fails to learn a generalized model for these five diverse sub-tasks (e.g. the difference on action space and the goal of each subtask). Players trained with five-model base curricula can generate some fundamental policies but lack tactical movements since the correlations between related sub-tasks are ignored. The weighted cascading curriculum training can make further improvements based on base curriculum training because the correlation between related sub-tasks is retained and the policy can be optimized over the whole task from the perspective of a single player. However, the coordination within one team remains a weakness. The ICTA model significantly outperforms other training approaches since the coordination performance can be essentially improved by using the coordination curriculum switcher. The results of the 300-day online evaluation show that ICTA reaches up to around 70% win-rate than human players during 3v3 PVPs despite that human players can purchase equipments to become stronger and they may discover the weakness of our models as time goes by. We can conclude that ICTA can be used as a general method for handling SPG like basketball.

ACKNOWLEDGMENTS

Hongyao Tang and Jianye Hao are supported by the National Natural Science Foundation of China (Grant Nos.: 61702362, U1836214).
REFERENCES


