

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

Extended Abstract

Hangtian Jia<sup>1</sup>, Chunxu Ren<sup>1</sup>, Yujing Hu<sup>1</sup>, Yingfeng Chen<sup>1+</sup>, Tangjie Lv<sup>1</sup>, Changjie Fan<sup>1</sup>

Hongyao Tang<sup>2</sup>, Jianye Hao<sup>2</sup>

<sup>1</sup>Netease Fuxi AI Lab, <sup>2</sup>Tianjin 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

### ACM Reference Format:

Hangtian Jia<sup>1</sup>, Chunxu Ren<sup>1</sup>, Yujing Hu<sup>1</sup>, Yingfeng Chen<sup>1+</sup>, Tangjie Lv<sup>1</sup>, Changjie Fan<sup>1</sup>, Hongyao Tang<sup>2</sup>, Jianye Hao<sup>2</sup>. 2020. Mastering Basketball with Deep Reinforcement Learning: An Integrated Curriculum Training Approach. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020*, IFAAMAS, 3 pages.

\*Equal contribution by the first two authors. + Corresponding author.

## 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].

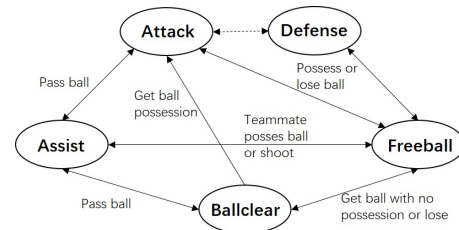


Figure 1: Correlations of Subtasks in Basketball

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

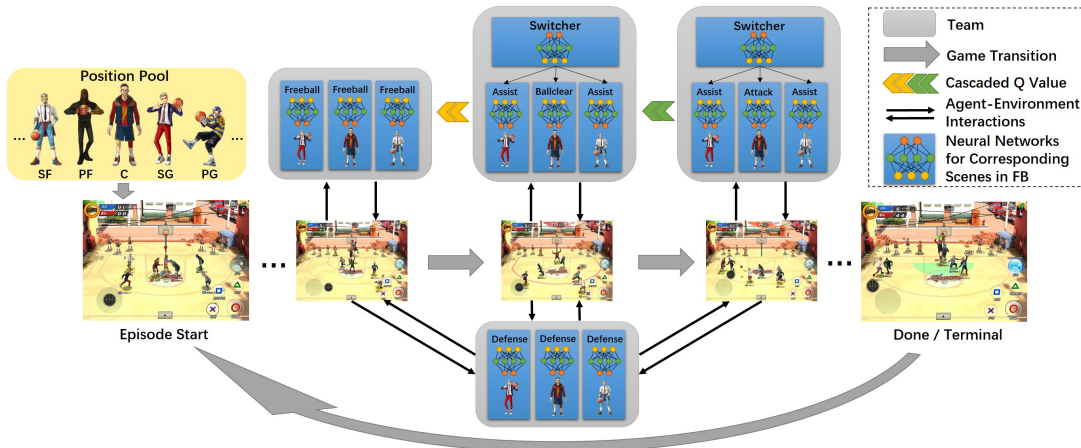


Figure 2: Scenes and Proposed Training Approaches in FB.

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  $\eta \in [0, 1]$ :

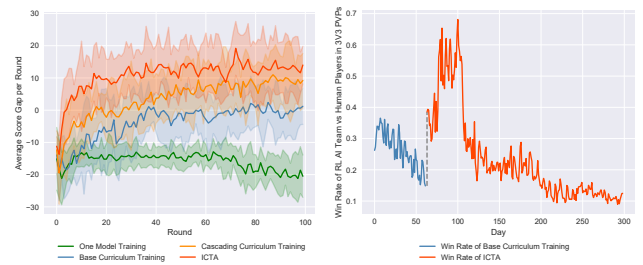
$$y_{\tau_i} = \begin{cases} r_i + \eta \gamma \max_{a'} Q_{\tau_j}^*(s'_{\tau_j}, a'_{\tau_j}; \theta_j), & \text{terminal s of } \tau_i \\ r_i + \gamma \max_{a'} Q_{\tau_i}^*(s'_{\tau_i}, a'_{\tau_i}; \theta_i), & \text{otherwise} \end{cases}$$

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<sup>1</sup> (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].

<sup>1</sup><https://github.com/FuxiRL/FeverBasketball>



(a) Training Process with Built-in AI (b) Win Rate during Online Evaluation

Figure 3: Performance of Our Models in FB.

## 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

- [1] Matthew M Botvinick, Yael Niv, and Andrew C Barto. 2009. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition* 113, 3 (2009), 262–280.
- [2] Johannes Heinrich and David Silver. 2016. Deep reinforcement learning from self-play in imperfect-information games. *arXiv preprint arXiv:1603.01121* (2016).
- [3] Matteo Hessel, Joseph Modayil, Hado Van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver. 2018. Rainbow: Combining improvements in deep reinforcement learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [4] Dan Horgan, John Quan, David Budden, Gabriel Barth-Maron, Matteo Hessel, Hado Van Hasselt, and David Silver. 2018. Distributed prioritized experience replay. *arXiv preprint arXiv:1803.00933* (2018).
- [5] Daniel R Jiang, Emmanuel Ekwedike, and Han Liu. 2018. Feedback-based tree search for reinforcement learning. *arXiv preprint arXiv:1805.05935* (2018).
- [6] Tejas D Kulkarni, Karthik Narasimhan, Ardavan Saeedi, and Josh Tenenbaum. 2016. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. In *Advances in neural information processing systems*. 3675–3683.
- [7] Karol Kurach, Anton Raichuk, Piotr Stańczyk, Michał Zajac, Olivier Bachem, Lasse Espeholt, Carlos Riquelme, Damien Vincent, Marcin Michalski, Olivier Bousquet, et al. 2019. Google Research Football: A Novel Reinforcement Learning Environment. *arXiv preprint arXiv:1907.11180* (2019).
- [8] Guillaume Lample and Devendra Singh Chaplot. 2017. Playing FPS games with deep reinforcement learning. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [9] Andrew Levy, George Konidaris, Robert Platt, and Kate Saenko. 2018. Learning Multi-Level Hierarchies with Hindsight. (2018).
- [10] Tianyu Liu, Zijie Zheng, Hongchang Li, Kaigui Bian, and Lingyang Song. 2019. Playing Card-Based RTS Games with Deep Reinforcement Learning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 4540–4546. <https://doi.org/10.24963/ijcai.2019/631>
- [11] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602* (2013).
- [12] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (2015), 529.
- [13] OpenAI. 2018. OpenAI Five. <https://blog.openai.com/openai-five/>. (2018).
- [14] Tianmin Shu, Caiming Xiong, and Richard Socher. 2017. Hierarchical and interpretable skill acquisition in multi-task reinforcement learning. *arXiv preprint arXiv:1712.07294* (2017).
- [15] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *nature* 529, 7587 (2016), 484.
- [16] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. 2017. Mastering the game of go without human knowledge. *Nature* 550, 7676 (2017), 354.
- [17] Chen Tessler, Shahar Givony, Tom Zahavy, Daniel J Mankowitz, and Shie Mannor. 2017. A deep hierarchical approach to lifelong learning in minecraft. In *Thirty-First AAAI Conference on Artificial Intelligence*.
- [18] Oriol Vinyals, Igor Babuschkin, Junyoung Chung, Michael Mathieu, Max Jaderberg, Wojciech M Czarnecki, Andrew Dudzik, Aja Huang, Petko Georgiev, Richard Powell, et al. 2019. AlphaStar: Mastering the real-time strategy game StarCraft II. *DeepMind Blog* (2019).