The CoachAI Badminton Environment: A Novel Reinforcement Learning Environment with Realistic Opponents (Student Abstract)

Kuang-Da Wang, Wei-Yao Wang, Yu-Tse Chen, Yu-Heng Lin, Wen-Chih Peng

National Yang Ming Chiao Tung University, Hsinchu, Taiwan
gdwang.cs10@nycu.edu.tw, sf1638.cs05@nctu.edu.tw, {s109550094.cs09, clementlin.mg08}@nycu.edu.tw, wcpeng@cs.nycu.edu.tw

Abstract

The growing demand for precise sports analysis has been explored to improve athlete performance in various sports (e.g., basketball, soccer). However, existing methods for different sports face challenges in validating strategies in environments due to simple rule-based opponents leading to performance gaps when deployed in real-world matches. In this paper, we propose the *CoachAI Badminton Environment*, a novel reinforcement learning (RL) environment with realistic opponents for badminton, which serves as a compelling example of a turn-based game. It supports researchers in exploring various RL algorithms with the badminton context by integrating state-of-the-art tactical-forecasting models and real badminton game records. The *Badminton Benchmarks* are proposed with multiple widely adopted RL algorithms to benchmark the performance of simulating matches against real players. To advance novel algorithms and developments in badminton analytics, we make our environment open-source, enabling researchers to simulate more complex badminton sports scenarios based on this foundation. Our code is available at https://github.com/wywyWang/CoachAI-Projects/tree/main/CoachAI%20Badminton%20Environment.

Introduction

In recent years, researchers have dedicated significant efforts to predicting and understanding various behaviors in sports (Wang 2023). However, the limited availability of environments and opponents that facilitate accurate comparisons and performance evaluations as realistic simulations hinder new algorithms from being validated in real-world games. To address this issue, Kurach et al. (2020) introduced an RL football environment that allows researchers to efficiently compare different algorithms, and provides opponent AI built-in bots in a physics-based simulator. Nonetheless, there are no existing environments for turn-based sports to provide realistic opponents due to their fundamental departure from previous setups, which typically encompassed only physics-based or rule-based randomness and opponents.

Therefore, our study focuses on a compelling example of a turn-based sport – badminton. We introduce the *CoachAI Badminton Environment*, which aims to revolve around a meticulously crafted realistic badminton RL environment with simulated opponents. However, at least two main challenges hinder the development of realistic opponents: **1) Transition Dynamic from Opponents**: In turn-based sports, opponents serve as transition dynamics, where their actions determine the RL agent's next state, necessitating comprehensive modeling. **2) Dynamic Strategies and Randomness**: Opponent behavior and randomness should adapt to the current situation to prevent excessive predictability, enabling RL agents trained in these environments to adapt to real-world scenarios. To overcome these challenges, we integrate two state-of-the-art models in badminton behavior predictions into our environment, providing a fine-grained range of actions: **ShuttleNet** (Wang et al. 2022), a transformer-based approach for *stroke forecasting*, and **DyMF** (Chang, Wang, and Peng 2023), a graph-based framework for *movement prediction*. These models have proven their effectiveness in simulating player behaviors from real-world records; therefore, our environment empowers researchers to train a diverse range of RL algorithms from realistic opponents.

Badminton Environment

Markov Decision Process (MDP)

Our environment adheres to singles badminton's real-world rules. In each set, the winner has to reach 21 points, with a two-point lead needed at 20-20; or, the first to 30 points wins. Each badminton game can be modeled as an Markov Decision Process (MDP) (Puterman 2014) with the following components: **State.** The player's state consists of the type of shot to receive, shuttlecock position, player position, and opponent position. Following the definition of Wang et al. (2022), there are 12 shot types in the environment: *receiving*, *short service*, *long service*, *net shot*, *clear*, *push/rush*, *smash*, *defensive shot*, *drive*, *lob*, *drop*, and *can't reach*. **Action.** We adopt the action space with the above 12 shot types. Both players take actions for each stroke consisting of the *landing position of the shuttlecock*, the *shot type*, and the *moving position to go to after returning the shuttlecock*. **Reward.** The reward follows badminton rules, where winning a rally earns the agent a point. This includes the opponent's shuttlecock landing out of bounds or when the shot type is *can't reach*. **Transition Dynamics.** In turn-based sports, the transition from the previous state to the current state is defined by the actions of the opponent.

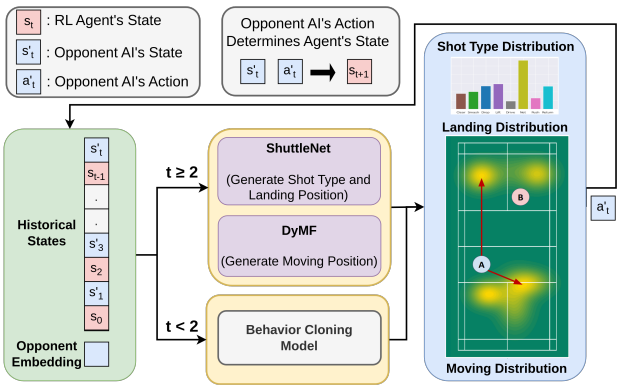


Figure 1: Illustration of the Opponent AI framework.

Opponent AI

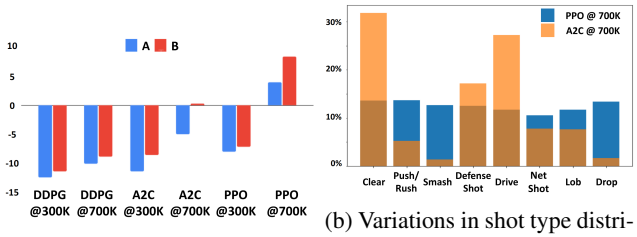
To tackle the first issue, Opponent AI is proposed as a transition dynamic to determine the opponent's action. As shown in Figure 1, the input comprises historical states, encompassing both RL agent and opponent AI experienced states, along with an Opponent Embedding indicating the current opponent. Depending on the historical states' length, they are directed to ShuttleNet and DyMF or the Behavior Cloning Model to generate opponent AI's actions.

The combination of ShuttleNet and DyMF. ShuttleNet predicts shot type and landing position from past states, while DyMF forecasts player movement from past states. Both models can be easily invoked with a state to produce a corresponding player's action since the input of these models is the same as a state. To enhance the accuracy of position prediction and the realism of randomness, we expand from a single bivariate Gaussian distribution to have both the landing and moving distributions follow weighted bivariate normal distributions. To validate our integrated approach, we employ the validation proposed by Wang et al. (2022) to assess the accuracy of predicting future strokes in the rally based on the first two states. We utilize mean absolute error (MAE) to measure the precision of predicting landing and moving locations, and cross-entropy (CE) for predicting shot types. As shown in Table 1, our integrated approach outperforms both the standalone ShuttleNet and DyMF across all metrics.

The Behavior Cloning Model. However, since both models require at least two historical states to predict future strokes, we introduce a Behavior Cloning (BC) model for the initial two strokes. We evaluate the BC model by predicting the first two strokes based on the rally's initial state. The BC model achieves a CE value of 1.0662 for shot type prediction, and MAE values of 1.3647 and 0.8770 for landing and moving positions, respectively. These results show that the BC model captures player decisions for the first two strokes when historical states are insufficient for ShuttleNet and DyMF. Integrating the BC model with ShuttleNet and DyMF enriches the environment with realistic opponents.

Model	Shot	Land	Move
ShuttleNet	2.4699	1.4276	-
DyMF	2.1094	-	1.6699
Ours (integration)	**1.8442**	**1.4060**	**1.6007**

Table 1: Quantitative results. Since ShuttleNet predicts shot type and landing position, and DyMF predicts shot type and moving position, the symbol − denotes unavailable results.



(a) Average score difference.

(b) Variations in shot type distributions.

Figure 2: Quantitative and qualitative results of the *Badminton Benchmarks*.

Badminton Benchmarks

We propose the *Badminton Benchmarks*, a predefined set of benchmark tasks to facilitate comparisons of algorithms against opponents in our environment, where the goal is to beat the opponent provided by our environment.

Experimental Setup. We benchmark three widely used RL algorithms: Deep Deterministic Policy Gradient (DDPG) (Lillicrap et al. 2016), Advantage Actor-Critic (A2C) (Mnih et al. 2016), and Proximal Policy Optimization (PPO) (Schulman et al. 2017) against two different opponent AIs, each representing a distinct real-world opponent. During the testing phase, the agents compete against their opponents in 20 sets, and we evaluate the average score difference.

Preliminary Results. Figure 2a displays results with two real-world players, representing different strengths. The opponent AI selection significantly influences the average score difference. A2C only manages to outperform Opponent B after 700K training steps, while it struggles to compete against Opponent A.

Case Study. We analyze A2C and PPO performance against Opponent A with 700K training steps, shown in Figure 2b. Comparatively, A2C primarily employs *Clear*, *Drive*, and *Defensive Shot*, indicating a conservative strategy. On the other hand, PPO exhibits a more even distribution of shot types, suggesting that diversifying shot type selection could be a key factor in defeating Opponent A. These results show that the *Badminton Benchmarks* offer applications for exploring winning strategies.

Acknowledgments

This work was supported by the Ministry of Science and Technology of Taiwan under Grants 112-2425-H-A49-001.

References

Chang, K.; Wang, W.; and Peng, W. 2023. Where Will Players Move Next? Dynamic Graphs and Hierarchical Fusion for Movement Forecasting in Badminton. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, 6998–7005. AAAI Press.

Kurach, K.; Raichuk, A.; Stanczyk, P.; Zajac, M.; Bachem, O.; Espeholt, L.; Riquelme, C.; Vincent, D.; Michalski, M.; Bousquet, O.; and Gelly, S. 2020. Google Research Football: A Novel Reinforcement Learning Environment. In *AAAI*, 4501–4510. AAAI Press.

Lillicrap, T. P.; Hunt, J. J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; and Wierstra, D. 2016. Continuous control with deep reinforcement learning. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Lillicrap, T. P.; Harley, T.; Silver, D.; and Kavukcuoglu, K. 2016. Asynchronous Methods for Deep Reinforcement Learning. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, 1928–1937. JMLR.org.

Puterman, M. L. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms. *CoRR*, abs/1707.06347.

Wang, K. 2023. Enhancing Badminton Player Performance via a Closed-Loop AI Approach: Imitation, Simulation, Optimization, and Execution. In Frommholz, I.; Hopfgartner, F.; Lee, M.; Oakes, M.; Lalmas, M.; Zhang, M.; and Santos, R. L. T., eds., *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, 5189–5192. ACM.

Wang, W.; Shuai, H.; Chang, K.; and Peng, W. 2022. ShuttleNet: Position-Aware Fusion of Rally Progress and Player Styles for Stroke Forecasting in Badminton. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, 4219–4227. AAAI Press.