RallyDiffuser: A Representation-Guided Diffusion Model Framework for Strategic Planning in Badminton

Extended Abstract

Bing-Zhi Ke
Department of Computer Science, National Yang Ming Chiao Tung University
Hsinchu, Taiwan
jklzxcvbnm1225.cs11@nycu.edu.tw

Kuang-Da Wang
Department of Computer Science, National Yang Ming Chiao Tung University
Hsinchu, Taiwan
gdwang.cs10@nycu.edu.tw

Wen-Chih Peng
Department of Computer Science, National Yang Ming Chiao Tung University
Hsinchu, Taiwan
wcpeng@cs.nycu.edu.tw

ABSTRACT

The rising interest in sports analysis has led to many studies from various perspectives, such as strategic insights and behavior prediction. In the rapid tactic nature of turn-based sports, badminton stands out as a compelling example of a game requiring players to make strategy-oriented decisions. Exiting planning works fail to capture its complex decision-making dynamics, particularly in balancing long-term strategy execution with immediate scoring opportunities in a turn-based setting. In this work, we propose RallyDiffuser, an innovative representation-guided diffusion model for strategic planning in badminton. We build a strategy latent space through representation learning that captures the variations in player strategies executed during rallies, and it identifies strategic anchors that guide agents in balancing long-term strategic objectives with short-term scoring opportunities. Our experiments demonstrate that RallyDiffuser outperforms existing planning methods, emerging as the only approach that achieves improved win rates across all strategies.

KEYWORDS

Offline Reinforcement Learning, Classifier-guided Diffusion Model, Badminton, Sport Analytics, Strategy Optimization

ACM Reference Format:
Bing-Zhi Ke, Kuang-Da Wang, and Wen-Chih Peng. 2025. RallyDiffuser: A Representation-Guided Diffusion Model Framework for Strategic Planning in Badminton: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

In recent years, with the widespread availability of player tracking data, the field of athlete performance enhancement has witnessed significant advancements, such as deep imitation learning in football [7] and Markov Decision Processes [10] in tennis Liu et al. [8] to simulate team dynamics, optimize play strategies, and improve win probabilities by analyzing complex decision-making patterns.

In this paper, we focus on badminton, a fast-paced turn-based sport. Due to the turn-based nature of the badminton game, a player's state is influenced by the opponent's actions. This interdependence requires that players switch their long-term badminton strategies to capitalize on short-term scoring opportunities arising from opponent mistakes, thus maximizing their chances of winning, which has not been addressed in prior works. This paper aims to maximize winning chances in badminton by identifying when to maintain long-term strategies and when to seize short-term scoring opportunities. This insight helps players and coaches refine their approach and improve their future performance.

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

2 BACKGROUND AND RELATED WORK

To generate the highest win-rate behaviors (i.e., the content of the rally) based on specified badminton strategies, we formulate it as a planning task and Use the most comprehensive publicly available badminton singles match dataset, ShuttleSet [12]. To utilize the dataset in planning work, we want the decision-making process in badminton games formulated in the Markov Decision Process (MDP) [10], defined by the tuple $\langle S, A, P_{\mathcal{T}}, R, \gamma \rangle$. One potential solution is the offline reinforcement learning (RL) model [6, 13], which has been successfully used for optimization from offline data across various domains, including treatment Optimization [9] and robot manipulation [2, 4]. Recent offline RL approaches using diffusion models [1, 5] address the issues of error accumulation and sparse rewards. They achieve this by focusing on trajectory planning with long-term reward accumulation.

Our method improves upon the diffusion-based method by introducing a novel representation-guided diffusion model framework to improve the win rate within the specific strategy. We propose a strategy latent space learned from the ratings of badminton and scoring strategies to tackle the strategy-opportunity trade-off and increases win rates by an average of 6.46% at the set level across different strategies.

3 METHOD

Figure 1.a illustrates an overview of the proposed RallyDiffuser framework, which consists of the following three stages. **Diffuser Training Stage:** Learn the player behavior across multimodal action distribution using the denoising diffusion process. We represent the current state-action pair with future states and actions into the two-dimensional array $\boldsymbol{\tau}$. Diffuser subsequently parameterizes a learned gradient ϵ_θ for the trajectory denoising procedure. The loss function is given by:

$$\mathcal{L}(\theta) = \mathbb{E}_{i,\epsilon,\boldsymbol{\tau}^0}\left[\|\epsilon - \epsilon_\theta(\boldsymbol{\tau}^i, C_t^r, i)\|^2\right], \quad (1)$$

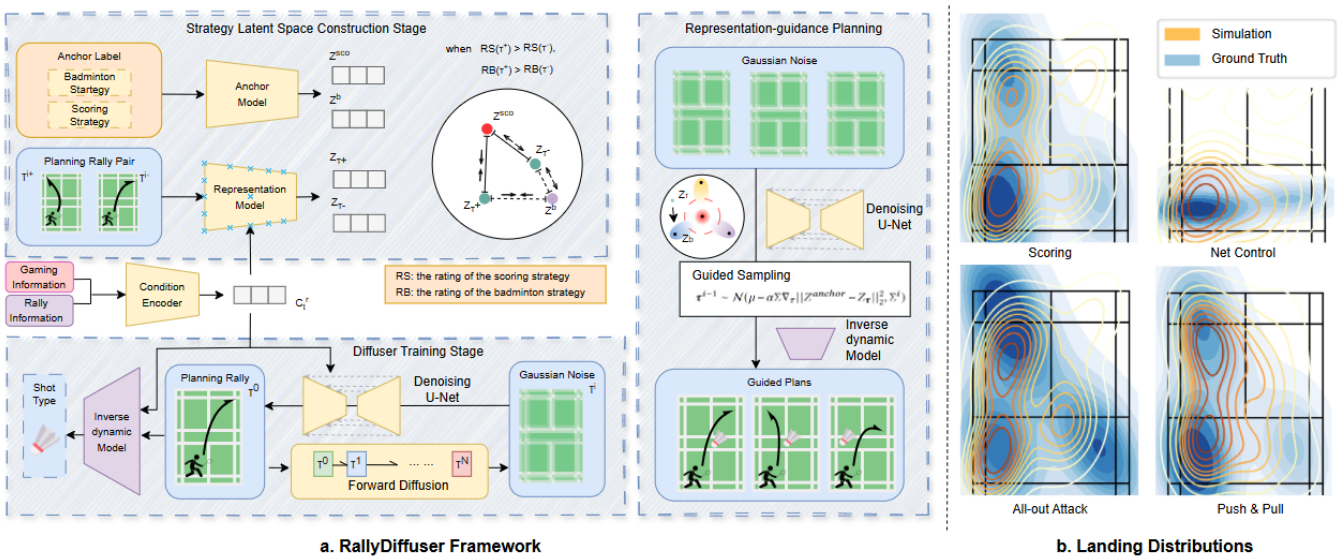

Figure 1: a. The framework of RallyDiffuser. b. The landing distributions guided by the different badminton strategies.

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ denotes the noise target, $i \sim \mathcal{U}\{1, 2, \ldots, N\}$ means the diffusion step, C_t^r is the embedd feature, and $\boldsymbol{\tau}^i$ is the trajectory $\boldsymbol{\tau}^0$ corrupted with noise ϵ.

Strategy Latent Space Construction Stage: Leverage representation learning with ratings for long-term badminton strategies (RB) and short-term scoring strategies (RS) to create a latent space where both can coexist. According to the two kind of rating, we can sample the pairs of $(\boldsymbol{\tau}^+, \boldsymbol{\tau}^-)$ required for each badminton strategy. For the specific anchor of the strategy, $\boldsymbol{\tau}$ with the higher rating of the strategy is viewed as the positive sample $\boldsymbol{\tau}^+$ while the other one is the negative sample $\boldsymbol{\tau}^-$. The loss $\mathcal{L}_{RB}$ and $\mathcal{L}_{RS}$ encourages the models respectively to map the representations with the high rating of the strategy $Z_{\boldsymbol{\tau}^+}$ and the anchor Z^b or Z^{sco} closer to each other, while the anchors in different badminton strategies keep away from each other:

$$\mathcal{L}_{RX} = \mathbb{E}\left[\max(d(Z_{\boldsymbol{\tau}^+}, Z^X) - d(Z_{\boldsymbol{\tau}^-}, Z^X) + \delta_{RX}, 0)\right], \quad (2)$$

where $d(\cdot)$ is the distance function, X means badminton strategy or scoring strategy, and δ_{RX} is a constant that separates the representation into different strategies.

Representation-Guidance Planning Stage: Utilize the anchor representations from the strategy latent space to guide decision-making, balancing long-term and short-term strategies during the game. With the trajectory representation Z_τ and the optimal anchors Z^{anchor} in strategy latent space, the condition sampling can be written as:

$$p_\theta(\boldsymbol{\tau}^{i-1} \mid \boldsymbol{\tau}^i, O_{1:T}) \approx \mathcal{N}(\mu - \alpha\Sigma\nabla_\tau ||Z^{anchor} - Z_{\boldsymbol{\tau}}||_2^2, \Sigma^i), \quad (3)$$

where $O_{1:T}$ means the expected event, α is a restricted scalar of the guidance. Benefiting from the strategy latent space, we select a distance threshold d. When $||Z^{sco} - Z_{\boldsymbol{\tau}}||_2 < d$, it considers that the player in the current state is likely to score the rally. At this point, the execution of the badminton strategy should be halted, and actions to score the rally should be undertaken instead. The flexibility helps RallyDiffuser catch the scoring moment within the long-term badminton strategy.

Table 1: Quantitative results. The objective is to improve the set-level win rate (SWR) under the limitation that the SCR has to be larger than the ImitationDiffuser. For each metric, the best result is highlighted in boldface, while the second-best result is underlined.

Model	All-out Attack w/ Scoring		Push and Pull w/ Scoring	
	SWR (↑)	SCR (↑)	SWR (↑)	SCR (↑)
ImitateDiffuser	48.06	36.82	48.06	40.95
IQL	45.54	-5.77	38.00	+19.97
DT	<u>70.83</u>	+31.12	71.13	-9.12
PEDA	36.00	+32.01	<u>44.79</u>	+5.88
DD	42.42	-1.81	37.76	+19.63
Diffuser	64.65	+3.33	74.75	-7.12
RallyDiffuser	**81.25**	+1.95	**57.29**	+7.33

3.1 Experiment and Conclusion

To demonstrate that RallyDiffuser improves the win rate under each badminton strategy, we compare it to the potential solutions, including IQL [6], DT [3], DD[1], Diffuser [5], and PEDA [14], conducted on the CoachAI badminton environment [11]. RallyDiffuser achieves great improvement for both the All-out Attack and Push & Pull strategies. On average, it reaches 68.05% in SWR, which respectively outperforms the second-best results of 6.46%. This approach maintains the long-term objective of badminton strategy execution while simultaneously addressing the short-term goal of scoring rallies. What's more, Figure 1.b shows the simulated distributions of the landing position, RallyDiffuser accurately capturing the landing distribution of all the strategies. Its capacity to provide strategic insights, aiding coaches and players in designing effective training and match plans. The quantitative evaluation demonstrates that RallyDiffuser outperforms existing planning methods and has the potential for sports analytics.

REFERENCES

[1] Anurag Ajay, Yilun Du, Abhi Gupta, Joshua Tenenbaum, Tommi Jaakkola, and Pulkit Agrawal. 2022. Is conditional generative modeling all you need for decision-making? *arXiv preprint arXiv:2211.15657* (2022).

[2] Yevgen Chebotar, Karol Hausman, Yao Lu, Ted Xiao, Dmitry Kalashnikov, Jacob Varley, Alex Irpan, Benjamin Eysenbach, Ryan Julian, Chelsea Finn, and Sergey Levine. 2021. Actionable Models: Unsupervised Offline Reinforcement Learning of Robotic Skills. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 1518–1528. http://proceedings.mlr.press/v139/chebotar21a.html

[3] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Misha Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. 2021. Decision transformer: Reinforcement learning via sequence modeling. *Advances in neural information processing systems* 34 (2021), 15084–15097.

[4] Nico Gürtler, Sebastian Blaes, Pavel Kolev, Felix Widmaier, Manuel Wuthrich, Stefan Bauer, Bernhard Schölkopf, and Georg Martius. 2023. Benchmarking Offline Reinforcement Learning on Real-Robot Hardware. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net. https://openreview.net/forum?id=3k5CUGDLNdd

[5] Michael Janner, Yilun Du, Joshua B Tenenbaum, and Sergey Levine. 2022. Planning with diffusion for flexible behavior synthesis. *arXiv preprint arXiv:2205.09991* (2022).

[6] Ilya Kostrikov, Ashvin Nair, and Sergey Levine. 2021. Offline reinforcement learning with implicit q-learning. *arXiv preprint arXiv:2110.06169* (2021).

[7] Hoang M Le, Peter Carr, Yisong Yue, and Patrick Lucey. 2017. Data-driven ghosting using deep imitation learning. MIT sloan sports analytics conference Boston, MA, USA.

[8] Zhaoyu Liu, Kan Jiang, Zhe Hou, Yun Lin, and Jin Song Dong. 2023. Insight Analysis for Tennis Strategy and Tactics. In *2023 IEEE International Conference on Data Mining (ICDM)*. IEEE, 1169–1174.

[9] Mila Nambiar, Supriyo Ghosh, Priscilla Ong, Yu En Chan, Yong Mong Bee, and Pavitra Krishnaswamy. 2023. Deep Offline Reinforcement Learning for Real-world Treatment Optimization Applications. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, Ambuj K. Singh, Yizhou Sun, Leman Akoglu, Dimitrios Gunopulos, Xifeng Yan, Ravi Kumar, Fatma Ozcan, and Jieping Ye (Eds.). ACM, 4673–4684. https://doi.org/10.1145/3580305.3599800

[10] Martin L Puterman. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

[11] Kuang-Da Wang, Yu-Tse Chen, Yu-Heng Lin, Wei-Yao Wang, and Wen-Chih Peng. 2024. The CoachAI Badminton Environment: Bridging the Gap between a Reinforcement Learning Environment and Real-World Badminton Games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 23844–23846.

[12] Wei-Yao Wang, Yung-Chang Huang, Tsi-Ui Ik, and Wen-Chih Peng. 2023. ShuttleSet: A Human-Annotated Stroke-Level Singles Dataset for Badminton Tactical Analysis. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5126–5136.

[13] Tianhe Yu, Aviral Kumar, Rafael Rafailov, Aravind Rajeswaran, Sergey Levine, and Chelsea Finn. 2021. Combo: Conservative offline model-based policy optimization. *Advances in neural information processing systems* 34 (2021), 28954–28967.

[14] Baiting Zhu, Meihua Dang, and Aditya Grover. 2023. Scaling Pareto-Efficient Decision Making via Offline Multi-Objective RL. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. https://openreview.net/forum?id=Ki4ocDm364