9th Asia-Pacific Conference on Plasma Physics, 21-26 Sep, 2025 at Fukuoka

Offline Reinforcement Learning by Decision Transformer for Tokamak Plasma Control

R. Yoneda[1], T. Kojima[1], Y. Shirasawa[1], M. Takahashi[1]
[1] NTT Space Environment and Energy Laboratories
e-mail (speaker): ryota.yoneda@ntt.com

Controlling hot and dense fusion plasma is a key technology to realize a fusion energy source. Approaches based on physics have been intensively investigated since the beginning of fusion plasma research. In recent years, along with the rapid development of computational power, machine learning or data-driven methods became attractive for the fields of plasma with complex and non-linear dynamics.

Devices that hold magnetically confined plasma such tokamaks and stellarators, the development of operational schemes remains challenging. If we state this as an optimization problem, reinforcement learning (RL) is one solution to control plasma utilizing the action inputs and observation outputs from the system [1,2]. RL usually requires interactions with environment so that an agent learns what is the best or optimal action to maximize the reward setting. In the field of interest, an agent interacts with environment, refers to online RL, may lead to failure of the experiments. This is particularly an issue for the larger devices such ITER where the damage of failure is not acceptable.

Offline RL (Fig.1(a)), on the other hand, does not require interactions with environments but only previously gained or generated data [3]. This technique seems to be appealing to controlling fusion plasma, but it comes with several downsides. Mainly because the entire offline data is fixed and not sufficient to provide a policy of decision making that maximizes the reward. To tackle this problem, one of the interesting approaches can be utilizing Decision Transformer [4]. Transformers achieved remarkable results in natural language processing represented by ChatGPT. Decision Transformer focuses on sequence modeling and conditional prediction, leveraging its ability to oversee long-range dependencies and sequence data.

As a RL environment, we employed TASK transport code [5] for preparing offline data of tokamak operation. To demonstrate offline RL learning, we set an action as ECH heating power, states as electron temperature T_e and density n_e based on Markov Decision Process (MDP). Reward increases as it approaches target T_e = 1.6 keV in the setup. During the discharge of 150 sec, reward setting becomes strict after 100 sec to validate that Decision Transformer can learn the requirement alterations during the operations. In this setup, Decision Transformer model was trained and tested. When evaluating the model, the inputs are only the states and it controls the next action. Also in this test, the trained model does not learn from the interactions from the environment.

The test results of model controls are shown in Fig.1(b). The model activates ECH control at 50 sec because during the plasma initiation, ECH has an insignificant effect on T_e and density n_e. After 100 sec when reward setting becomes strict, we confirm that the model controls T_e to get close to the target to keep reward high. We investigated how dataset size affects the model behavior. When increasing the size of train data, it adjusts ECH power to reach the target earlier after control starts.

In the presentation, we will discuss in detail how decision transformer models behave with the effect of data size, more complicated plasma parameter optimization and time-sensitive control during the operation.

References:
[1] Seo, Jaemin et al. *Nature* vol. 626,8000 (2024): 746-751.
[2] Degrave, Jonas et al. *Nature* vol. 602,7897 (2022): 414-419.
[3] Levine, Sergey, et al. *arXiv preprint* arXiv:2005.01643 (2020).
[4] Chen, Lili, et al. *Advances in neural information processing systems* 34 (2021): 15084-15097
[5] A. Fukuyama, et al., *Plasma Phys. Control. Fusion* **37**, 611 (1995)

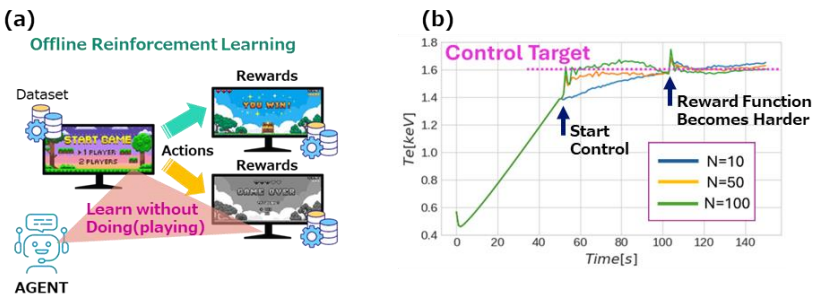

Figure 1: (a)Offline RL overview, an agent learns the optimum policy without environment interactions. (b)T_e evolution with model control by Decision Transformer. Tested control with dataset number N=10, 50 and 100 shots. Target value as T_e = 1.6keV (pink dot line).