Investigation of q-profile and normalized beta control in JT-60SA using reinforcement learning

T. Wakatsuki[1], T. Suzuki[1], N. Hayashi[1], M. Yoshida[1]

[1] *National Institutes for Quantum Science and Technology, Naka, Japan*

1. Introduction

Real-time control of a safety factor (q) profile and a normalized beta (β_N) has been studied in various tokamaks because they strongly affect the confinement performance and MHD stability of fusion plasmas. Especially for advanced tokamak scenarios, it is very important to tailor the q profile to have a weak or negative magnetic shear in the inner half of the plasma since the anomalous transport can be reduced and, in some cases, the internal transport barrier (ITB) can be produced. In such plasmas, a large amount of plasma current is driven by self-generated bootstrap current, and therefore, the q profile and β_N tend to be strongly linked. In addition, the confinement characteristics can change during the discharges. Due to these difficulties, the design of the q profile and β_N controller for the advanced tokamak is very challenging.

In this work, an integrated q profile and β_N control for JT-60SA is developed using reinforcement learning. A neural network (NN) based controller is trained through trial-and-errors that are performed in the simulation using an integrated transport code more than one million times. The integrated transport code RAPTOR [1] is used because it enables us to run rapid predictive simulations. In our previous work, it was shown that the NN controller trained for the control of the ion temperature gradient using reinforcement learning could be robustly used even if the controlled parameter had a wide range of response characteristics [2]. In this work, an integrated q profile and β_N controller that can be used in advanced plasmas with a wide range of ITB strength has been developed. It should not be suitable to set the q profile itself as a target value since the q profiles that are feasible with a finite set of actuators will be different if confinement characteristics are different. Therefore, the target of the minimum value of the q profile (q_{min}) is set as a target value instead. With this target, the constraint is relaxed while the control of the plasma performance, such as stability and confinement characteristics is maintained.

Since the experiments in JT-60SA have not yet started, the validity of the trained NN controller is checked using the other integrated code TOPICS [3]. TOPICS is often used for predictive simulations for JT-60SA. Although both RAPTOR and TOPICS solve 1D transport

equations, they use different physics models. For example, they use different heating and current drive models and anomalous transport models. Therefore, if the trained NN is tested using TOPICS, its validity for the plasmas whose response characteristics are different from the plasmas used in the training can be checked. If it is valid, it is expected to be applicable to the real experiment.

2. Training using RAPTOR code

The controller is trained to control q_{min} and β_N by neutral beams (NBs) and electron cyclotron (EC) waves. JT-60SA will be equipped with on-axis and off-axis negative ion-based neutral beams (on-axis NNB and off-axis NNB), and three groups of positive ion-based neutral beams which will be injected in co-current tangential (co-PNB), counter-current tangential (ctr-PNB) and perpendicularly (perp-PNB). The input parameters of the NN are defined as in the state $\mathbf{s}_i$ = [t_i, $\mathbf{q}(t_i, \rho)$, $\mathbf{q}(t_{i-1}, \rho)$, $\mathbf{T_e}(t_i, \rho)$, $\mathbf{P_{act}}(t_i)$, $\beta_N(t_i)$, $\overline{\beta_N}(t_i)$, $\beta_N^{limit}(t_i)$, $\theta_{Flattop}$], where $\mathbf{P}_{act}(t_i)$ = [P_{EC}, $P_{on\text{-}NNB}$, $P_{off\text{-}NNB}$, P_{co}, P_{ctr}, P_{perp}] is the heating power of EC and each group of NBs, $\overline{\beta_N}(t_i)$=$0.8\overline{\beta_N}(t_{i-1})$+$0.2\beta_N(t_{i-1})$ is the smoothed value of the past β_N data, and $\mathbf{P}_{act}(t_{i+1})$ is the output parameter of the NN. The smoothed β_N is included in the state since it can be used as a control target to reduce the oscillation in β_N control. The parameters included in the state are chosen such that the NN can observe the present response characteristics of the q profile and β_N to the heating and current drive by EC and NBs. In reinforcement learning, the interaction of the controller and the plasma is modelled as a transition of state $\mathbf{s}_i$ to the state at the next control time step $\mathbf{s}_{i+1}$ due to an output of the controller, $\mathbf{P}_{act}(t_i)$. The NN is trained using a parameter called a reward which evaluates the control result associated with the state transition at each time step. The reinforcement learning algorithm tries to maximize a sum of rewards obtained in a series of time-dependent simulations.

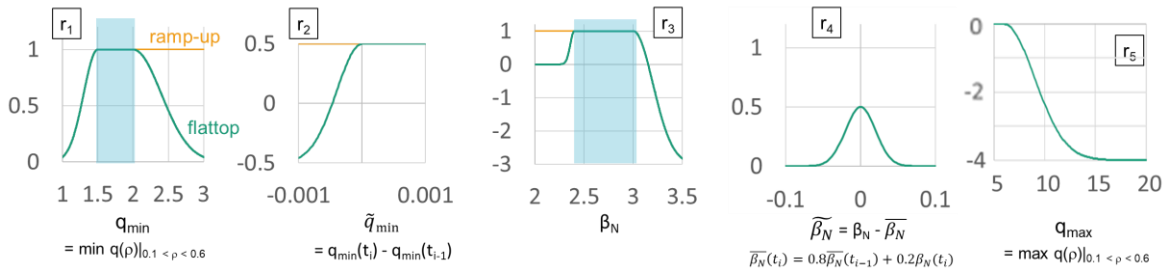

Fig. 1 Definition of five terms of rewards. A reward at each time step is a sum of these terms. Rewards of r_1 and r_3 are added to encourage q_{min} and β_N to be controlled in each target range shown in cyan. Rewards of r_2 and r_4 are given to encourage to find a stationary phase. The last term r_5 is added to avoid strongly reversed shear plasmas or current hole plasmas. The green curves and yellow curves denote the rewards in flat-top phase and ramp-up phase, respectively.

An ad-hoc anomalous transport model which describes the confinement improvement due to low magnetic shear is used. To train against plasmas with a wide range of confinement characteristics, the parameters used in this model are randomly changed at the beginning of each time-dependent simulation.

The design of the reward is a key to realize the control that fulfills the requirements. The control targets are not defined as specific values for q_{min} and β_N but as ranges of those values. This is because there is no set of q_{min} and β_N values that can be realized for all the plasmas used in the training. To realize advanced plasmas, the lower limit of β_N target is set to 2.4 and that of target q_{min} is set to 1.5. In addition, since higher q_{min} and β_N might cause confinement degradation and/or MHD instabilities, the upper limit of β_N target is set to 3.0 and that of target q_{min} is set to 2.0. Based on those targets, the reward is defined as shown in Fig. 1.

In the time-dependent simulation in training, the current ramp-up from 0.9 MA to 1.9 MA for 3.5 s followed by the current flat-top that lasts for 45 s in JT-60SA are simulated. The control time step is 100 ms, therefore, each simulation consists of 485-time steps of simulation. After more than two million time steps of training, the average of the sum of rewards in one time-dependent simulation exceeds 1000. Using the trained NN, q_{min} and β_N can be stably controlled

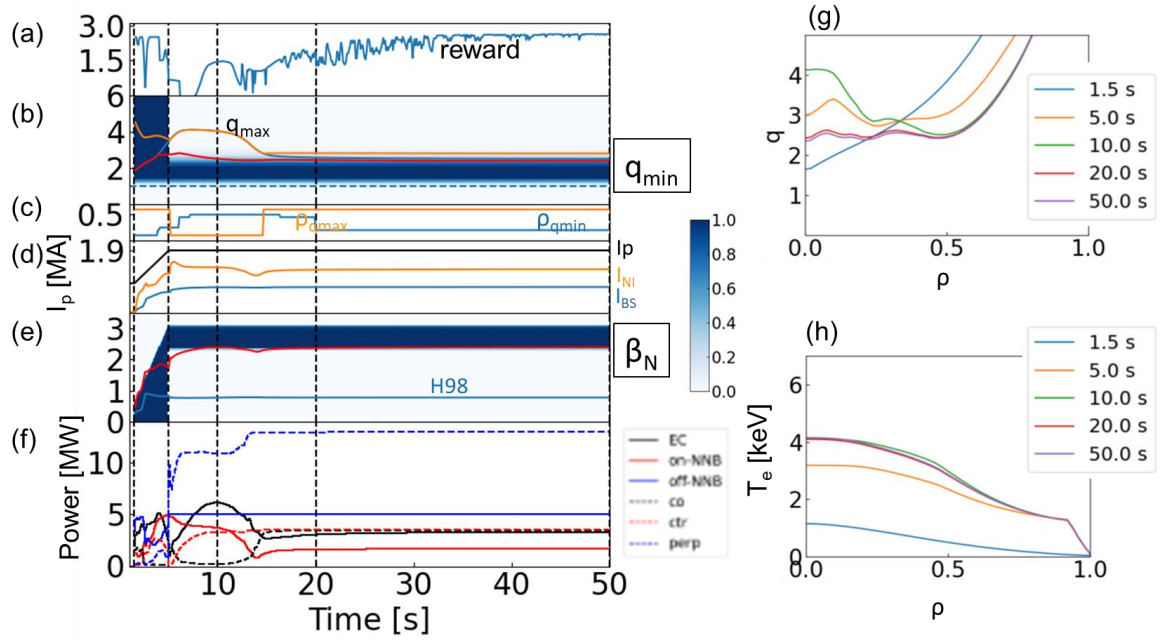

Fig. 2 Control results using the trained NN in TOPICS. Bohm/gyro-Bohm model is used. Left column shows the time evolution of (a) a reward, (b) q_{min} and q_{max}, (c) ρ_{qmin} and ρ_{qmax}, (d) I_p, I_{BS} and I_{NI}, (e) β_N and H98, and (f)input powers of EC and NBs. The hatched regions in (b) and (e) show the range of rewards associated with q_{min} and β_N, respectively. Right column shows the snapshots of (g) q profiles and (h) T_e profiles at the times shown by dashed lines in left column.

in target ranges for most cases simulated in RAPTOR. Nonetheless, the control for plasmas

with strong ITB is challenging. In such plasmas, q_{min} and β_N are kept within the target ranges but those values are oscillated in time scale more than 20 s.

3. Validation using TOPICS code

The trained NN is tested in the simulation using the integrated code TOPICS. In TOPICS, CDBM model or Bohm/gyro-Bohm model are used to simulate strong ITB plasmas or weak ITB plasmas. As shown in Fig. 2, q_{min} and β_N are stably controlled for plasmas with Bohm/gyro-Bohm model. Although q_{min} is 2.2 and slightly higher than the target range, for weak ITB plasmas, the trained NN can realize stable control of q_{min} and β_N. Note that the models used in TOPICS are different from those used in the training. Another simulation with CDBM shows a limitation of q_{min} control in the strong ITB plasmas. Although both q_{min} and β_N are stably controlled, q_{min} is higher than the target range.

4. Summary

A system for q profile and β_N control in JT-60SA has been developed using reinforcement learning. This system controls q_{min} and β_N in target ranges that correspond to advanced plasmas. This system is trained in more than two million times trials in RAPTOR simulations. In training, the model parameters that determine confinement property are randomly changed shot-by-shot. As a result of this randomization, the trained system realizes a stable control of q_{min} and β_N for weak ITB plasmas. This is confirmed by the simulation using TOPICS. Even though there are many differences in physics models and assumptions between RAPTOR and TOPICS, the trained system can achieve the stable control of q_{min} and β_N. This is an encouraging result for the application of the trained system to real experiments. However, at the same time, the issues in the development of the control system using reinforcement learning are found. In this work, the NN is trained for plasmas with a wide range of confinement characteristics. This setup makes it difficult to set a specified control target and it is shown difficult to achieve good control in challenging circumstances such as control in strong ITB plasmas. It will be required to train against plasmas that have confinement characteristics that are relevant to the target plasma scenarios.

[1] F. Felici et. al. Plasma Physics and Controlled Fusion 54(2), 2012, 025002

[2] T. Wakatsuki et. al. Nucl. Fusion 61 (2021) 046036

[3] N. Hayashi and JT-60 Team Phys. Plasmas 17, 2010, 056112