Mongolia microgrid energy storage

Kuo Wang, Zhanqiang Zhang, Keqilao Meng, Pengbing Lei, Rui Wang, Wenlu Yang, Zhihua Lin; Optimal energy scheduling for microgrid based on GAIL with Wasserstein distance. AIP Advances 1 August 2024; 14 (8): 085013. https://doi /10.1063/5.0207444

This paper is organized as follows: in Sec. II, the microgrid model is described. Section III outlines the energy optimal scheduling problem as a Markov decision process. Section IV presents the proposed methodology for solving the energy optimal scheduling problem. Our case study and results are discussed in Sec. V. Finally, our conclusions are summarized in Sec. VI.

Microgrids typically consist of renewable energy generation units, traditional fossil energy generation units, energy storage devices, and user loads. The microgrid system model is illustrated in Fig. 1. In the grid-connected mode, the microgrid interacts with the main grid (MG), and according to the power generated by the distributed power sources and the power used by the power loads, it carries out a reasonable power distribution, ultimately achieving optimal energy scheduling.

This paper proposes a WGAIL algorithm for optimal energy scheduling in microgrids, based on the constructed MDP problem. The algorithm is structured by expert policy data, a generator network, and a discriminator network. The policy network of the agent is updated using a reinforcement learning algorithm. After iterative updates with feedback from the discriminator, the optimal decision for the energy scheduling problem in the LDR scenario is finally obtained. The GAIL method, combining the Wasserstein distance and PPO algorithm, is shown in Algorithm 1.ALGORITHM 1.Generative adversarial imitation learning with Wasserstein distance.

Structure of the WGAIL algorithmic framework.

Imitation learning requires fitting expert datasets, and the quality of the expert data determines how well the agent learns. In this paper, an expert data collection method is designed to train in a microgrid environment using the PPO algorithm. The trained model serves as the expert policy model, and state-action expert data samples are collected through this model. The process of capturing expert experience is illustrated in Fig. 3.

Capturing the experience of experts.

Effect of advantage functions on strategy updating.

The PPO algorithm is a strategy gradient method based on the actor-critic framework, which contains two types of neural networks: the policy network (actor) and the value network (critic). The actor network takes the environment state as input and outputs the probability of each action. The critic network evaluates the current state and outputs an evaluation value. Generally, both networks share a common feature extraction layer.

In this paper, the state space is three-dimensional and the action space is four-dimensional, and we construct the neural network based on this setup. The parameters of the strategy network (actor) and the value network (critic) are shown in Tables I and II.

Actor network structure and parameterization.

Critic network structure and parameterization.