Optimal dispatch of Integrated Energy Systems by Fusing Graph Neural Network and Deep Reinforcement Learning

Qichuan Liu[1], **Weining Shi**[1], **Hongxu Zhang**[1],

[1]School of Informatics, Xiamen University
{30920231154354, 30920230157371, 30920231154367}@stu.xmu.edu.cn

Abstract

The optimal dispatch of Integrated Energy System (IES) plays a crucial role in achieving the objectives of multi-energy complementarity, energy conservation, and emission reduction. In recent years, with the development of artificial intelligence technology, some researchers have started to leverage the exploratory capabilities of Deep Reinforcement Learning (DRL) to address the optimization challenge faced by energy systems. Due to the expression of the system state as a vector and its subsequent use for training purposes, the intrinsic connection between nodes within the system is disregarded. As a result, these approaches inherently possess limitations in terms of training efficiency and exploration ability. We proposes a DRL model based on Distributional Soft Actor-Critic (DSAC) of Graph Neural Network (GNN) architecture, serving as an alternative to Multi-Layer Perceptron (MLP) architecture. More effective exploration and learning can be achieved by modeling the IES as graph structure data and feeding it into the model. Numerical simulations and comparative experiments confirm the method's advantages in terms of training efficiency and optimization results.

Introduction

The optimal dispatch of Integrated Energy Systems (IES) in Energy Internet is one of the core issues in multi-energy flow analysis, and it plays an important role in energy conservation, emission reduction, and the full utilization of energy resources (Li et al. 2020). Currently, research in this field mainly focuses on optimal power flow solutions in Electricity-Heat or Electricity-Gas energy systems (Tang et al. 2020; Ge et al. 2020), and in recent years, there have been increasing efforts in modeling and solving the Electricity-Heat-Gas systems (Yang et al. 2020). However, most of these solution approaches are centered around approximate or nonlinear solutions, inevitably facing challenges such as high algorithm complexity and the need to resolve the system state changes, making it difficult to achieve fast response in large-scale systems and unable to guarantee the attainment of global optimization. Additionally, the randomness of power output from new energy power stations has brought new difficulties to the optimization problem (Li et al. 2021). How to fully utilize new energy to reduce the use of traditional energy, achieve cost reduction, energy conservation, peak shaving, and valley filling in power grid operation has become an urgent problem that needs to be addressed (Zhou, Liu, and Liu 2014).

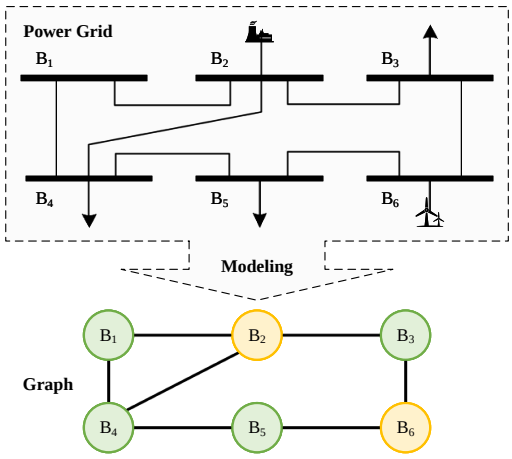

Figure 1: The simple power grid comprises 6 busbars, a generator, wind-power station and loads. Additionally, this system contains valuable topological information. When considering the busbars as nodes and the transmission lines as edges, we can easily model the power gird as a graph.

In recent years, Reinforcement Learning (RL) has made remarkable progress in solving combinatorial optimization problems (Mazyavkina et al. 2021), and has been validated in the optimal dispatch of Electricity-Heat integrated energy systems (Liu et al. 2021; Yang, Huang, and Chen 2021) and Electricity-Gas integrated energy systems (Qiao et al. 2021). However, these methods directly input the system state represented as a vector for training, ignoring the topological connectivity structure of the system, which inevitably results in the limitation of missing hidden physical correlations. Graph Neural Network (GNN), as a popular research area recently, can effectively utilize the topological information of systems to depict complex nonlinear relationships between nodes (Wu et al. 2021). By utilizing adjacency matrix to depict the connectivity relationships between nodes and realize the information propagation between nodes, GNN has effectively explored the complex non-Euclidean relationships

between nodes. In reactive power optimization problems in power systems, GNN has achieved better accuracy and robustness compared to traditional Multi-Layer Perceptron (MLP) (Liao et al. 2021).

For the optimal dispatch problem of IES that include new energy power stations, we proposes a Deep Reinforcement Learning (DRL) model based on GNN, which replaces the traditional fully connected neural network architecture. By modeling the Electricity-Heat IES as a graph-structured dataset, this model achieves more effective exploration and learning. The main contributions of this paper are as follows:

- We propose designing a GNN based on the physical topology of the IES. This approach allows us to fully explore potential links and correlations.
- We propose a novel RL model based on GNN and Distributional Soft Actor-Critic (DSAC). Our model achieves superior performance compared to existing RL-based solutions for optimal dispatch problems.
- We comprehensively compare the trend of Energy Internet metrics in the arithmetic example and analyze the steady-state factors of IES.

Related Work

Deep Reinforcement Learning. DRL are mainly divided into two categories: value-based algorithms (Liu, Gao, and Luo 2019) and policy gradient algorithms (Liu et al. 2018). Deep Q-Network (DQN), uses the experience playback mechanism to store the experience data of the intelligent body interacting with the environment online into the experience pool, and randomly samples the data in the experience pool in small batches during training to break the correlation between the data (Mnih et al. 2013). While the policy gradient algorithm directly employs the policy network to search for actions, specifically suitable for continuous action scenarios, the Deep Deterministic Policy Gradient (DDPG) adopts a deterministic approach to sampling actions to further enhance the generality of the algorithm (Lillicrap et al. 2015). In recent years, there has been significant attention given to the Actor-Critic (AC) structure, which combines both value-based algorithms and policy gradient algorithms. In this structure, the actor selects actions using the strategy gradient method, and the value function gives the score. The Soft Actor-Critic (SAC) algorithm adds the idea of entropy to the objective function (Haarnoja et al. 2018).

Graph Neural Networks. Graph Neural Network (GNN) aiming to extract and discover patterns within graph, which fulfills the requirements of graph representation learning tasks, including clustering, classification, prediction, generation, and more. The origins of GNN can be traced back to as early as 2005, (Gori, Monfardini, and Scarselli 2005) was first propose the concept of GNN, which use Recurrent Neural Network (RNN) to deal with undirected graphs, directed graphs and cyclic graphs. Afterwards, the GNN algorithms of this model were further inherited and enhanced (Micheli 2009). Afterwords, (Bruna et al. 2013) suggests employing Convolutional Neural Network (CNN) on graph. The Graph Convolutional Network (GCN) is introduced by cleverly adapting the convolution operator. GCN realizes the translation invariance, local perception and weight sharing of CNN on graphs (Bronstein et al. 2017). In contrast to GCN, which equally aggregates neighbor information, Graph Attention Networks (GAT) incorporates an attention mechanism to learn the varying importance of each neighbor (Veličković et al. 2017).

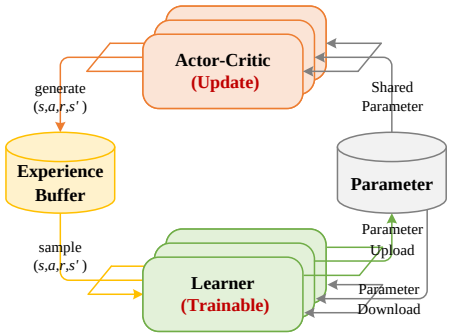

Figure 2: The architecture of DSAC. The Experience Buffer, Actor-Critics, and Learners are distributed across multiple workers, and Parameter communication between the different neural networks is asynchronous.

Integrated Energy Systems. Current research on The optimal dispatch of IES is roughly divided into three categories: the first category of research solely focuses on specific regions, which does not consider overall planning of multi-energy networks (Salimi et al. 2015); the second category considers not only various regions, but also the overall planning of multi-energy networks, which is typically employed in the model of direct current within the grid. (Huang et al. 2016); The third category can be summarized as the study of joint optimal problems of power grid, power supply and gas system, which focuses on the coupling of power grid and gas system.The demand side only considers the electricity and gas loads (Chaudry et al. 2014). Additionally, there are studies that utilize relaxation algorithms or linearization methods (Qiu et al. 2015) to simplify the optimal model for the gas system.

Proposed Solution

Distributional soft actor-critic (DSAC) algorithm, shown in Figure 2, is an off-policy RL method for continuous control setting, to improve the policy performance by mitigating Q-value overestimations (Duan et al. 2021). Considering the properties of IES is naturally a graph, we propose the DSAC algorithm that combines the GNN-based actor strategy.

Problem Formulation

An Electric-Heat-Gas Integrated Energy System (IES), including Power System, Thermal System, Natural Gas System and Coupling System, can be denoted as a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of edges.

State Space. For each system, a portion of its state space (i.e., node features and edge features) needs to be obtained from the power equations.

Power System State Space. For nodes $\mathcal{V}^{\text{elec}} \subset \mathcal{V}$ and edges $\mathcal{E}^{\text{elec}} \subset \mathcal{E}$ in the power system. For each node, its feature $\boldsymbol{v}_i^{\text{elec}} = [P_{i,t_0}, P_{i,t_1}, \ldots, P_{i,T}] \in \mathbb{R}^T$ is denoted as the electrical load P_i from t_0 to T. For each edge, its feature $\boldsymbol{e}_{ij}^{\text{elec}} = [G_{ij}, B_{ij}] \in \mathbb{R}^2$ is denoted as the concatenate of the conductance G_{ij} between node i and node j and the susceptance B_{ij} between node i and node j.

Thermal System State Space. For nodes $\mathcal{V}^{\text{heat}} \subset \mathcal{V}$ and edges $\mathcal{E}^{\text{heat}} \subset \mathcal{E}$ in the thermal system. For each node, its feature $\boldsymbol{v}_i^{\text{heat}} = [H_{i,t_0}, H_{i,t_1}, \ldots, H_{i,T}] \in \mathbb{R}^T$ is denoted as the heat load H_i from t_0 to T. For each edge, its feature $\boldsymbol{e}_{ij}^{\text{heat}} = \left[L_{ij}^{\text{heat}}, m_{ij}\right] \in \mathbb{R}^2$ is denoted as the concatenate of the length L_{ij}^{heat} of branch of pipeline ij and the mass flow rate m_{ij} between node i and node j.

Natural Gas System State Space. For nodes $\mathcal{V}^{\text{gas}} \subset \mathcal{V}$ and edges $\mathcal{E}^{\text{gas}} \subset \mathcal{E}$ in the natural gas system. For each node, its feature $\boldsymbol{v}_i^{\text{gas}} = [f_{i,t_0}, f_{i,t_1}, \ldots, f_{i,T}] \in \mathbb{R}^T$ is denoted as the gas load f_i from t_0 to T. For each edge, its feature $\boldsymbol{e}_{ij}^{\text{gas}} = \left[L_{ij}^{\text{gas}}, \kappa_{ij}\right] \in \mathbb{R}^2$ is denoted as the concatenate of the the length L_{ij}^{gas} of pipeline ij and the pipeline constant κ_{ij}.

Coupling System. In this paper, combined heat and power generation (CHP) is considered to meet the load demand of the three systems. The state space of CHP can be represented by the following linear programming:

$$\begin{aligned} \min\{&P_{\min}^{\text{CHP}} - \alpha_3 H^{\text{CHP}}, \alpha_1 + \alpha_2 H^{\text{CHP}} \\ &\leq P^{\text{CHP}} \leq P_{\max}^{\text{CHP}} - \alpha_3 H^{\text{CHP}}\} \end{aligned} \tag{1}$$

where P^{CHP} and H^{CHP} represent the electrical output and thermal output of the CHP, while $P_{\min}^{\text{CHP}}$ and $P_{\max}^{\text{CHP}}$ indicate the lower and upper limits of the electrical output. Similarly, $H_{\min}^{\text{CHP}}$ and $H_{\max}^{\text{CHP}}$ represent the lower and upper limits of the thermal output. Additionally, α_1, α_2, and α_3 are coefficients used for calculating the polygon area.

Action Space. Action space $\mathbf{A}$ has concrete physical significance. $\mathbf{A} = \{P^{\text{G}}, P^{\text{CHP}}, H^{\text{CHP}}, \alpha^{\text{W}}, f\}$, including the active power output of thermal power stations P^{G}, the electrical and thermal power output of CHP P^{CHP} and H^{CHP}, the wind power absorption coefficient α^{W}, and the gas supply volume of natural gas supply stations f.

Reward Function. The objective of the optimal dispatch task in IES is to minimize operating costs while ensuring constraint satisfaction.

Operating Costs. Operating costs include thermal power station operating costs, CHP operating costs and natural gas costs. First costs are calculated as follows:

$$F_1 = \sum_{t=t_0}^{T} \sum_{i=1}^{|N_{\text{P}}|} (\alpha_2 P_{i,t}^2 + \alpha_1 P_{i,t} + \alpha_0) \tag{2}$$

where T represents the total operating time. $|N_{\text{P}}|$ denotes the quantity of thermal power stations. $P_{i,t}$ signifies the active output of thermal power station i during time t. $\alpha_0, \alpha_1, \alpha_2$ represents the parameters of the consumption characteristic curve for the thermal power unit.

CHP operating costs are calculated as follows:

$$\begin{aligned} F_2 = \sum_{t=t_0}^{T} \sum_{i=1}^{|N_{\text{CHP}}|} &(\mu_0 + \mu_1 H_{i,t}^{\text{CHP}} + \mu_2 P_{i,t}^{\text{CHP}} + \\ &\mu_3 (H_{i,t}^{\text{CHP}})^2 + \mu_4 (P_{i,t}^{\text{CHP}})^2 + \mu_5 H_{i,t}^{\text{CHP}} P_{i,t}^{\text{CHP}}) \end{aligned} \tag{3}$$

where $|N_{\text{CHP}}|$ represents the quantity of CHP. P^{CHP}, H^{CHP} signify the electrical output and heat output, respectively, of the CHP during time t. μ represents the parameters of the consumption characteristic curve for the CHP.

natural gas costs are calculated as follows:

$$F_3 = \sum_{t=1}^{T} \sum_{i=1}^{|N_{\text{G}}|} C_{\text{gas}} f_{i,t} \tag{4}$$

where $|N_{\text{G}}|$ represents the quantity of gas supply stations. C_{gas} denotes the price per unit of natural gas. $f_{i,t}$ signifies the volume of gas supplied of each station during time t.

Physical Constraints. The stable state of IES requires compliance with safety constraints, including node voltage, line transmission power and so on. In addition, the IES needs to satisfy the ramping constraints, including electric power ramping and thermal ramping.

$$r = -(F_1 + F_2 + F_3 + \sum_{i=1}^{9} \lambda_i |\cdot|) \tag{5}$$

where λ_i signifies the penalty factor, while $|\cdot|$ denotes the safety constraints and ramping constraints. To ensure that the training results satisfy the constraints, generally set a larger penalty factor. When the constraints are satisfied, the penalty term equals 0.

In conclusion, the reward function encompasses the expenses associated with system operation as well as the penalties incurred for violation of constraints.

Distributional Soft Actor-Critic Algorithm

DSAC ensures the randomness of strategy learning and expands the exploration range as much as possible to prevent getting stuck in a local optimal solution. The goals of the DSAC algorithm are as follows:

$$G_\pi = \mathop{\mathrm{E}}_{\substack{(s_{i\geq t}, a_{i\geq t}) \sim \rho_\pi \\ r_{i\geq t} \sim R(\cdot|s_i, a_i)}} \left[\sum_{i=t}^{\infty} \gamma^{i-t} [r_i + \alpha \, \mathrm{H}(\pi(\cdot|s_t))] \right] \tag{6}$$

where $\mathrm{H}(\pi(\cdot|s)) = \mathrm{E}_{a\sim\pi(\cdot|s)}[-\log \pi(a|s)]$ represents the entropy value of the strategy $\pi(a|s)$ in the state s.

To evaluate the strategy π , define a soft Q-value function and use the Bellman operator T^π:

$$\begin{aligned} \mathrm{T}^\pi Q^\pi(s,a) =& \mathrm{E}[r] + \gamma \, \mathrm{E}[Q^\pi(s_{t+1}, a_{t+1}) \\ &- \alpha \log \pi(a_{t+1}|s_{t+1})] \end{aligned} \tag{7}$$

The objective of policy improvement is to identify a new policy π_{new} that outperforms the current policy, thus yielding a higher expectation of return.

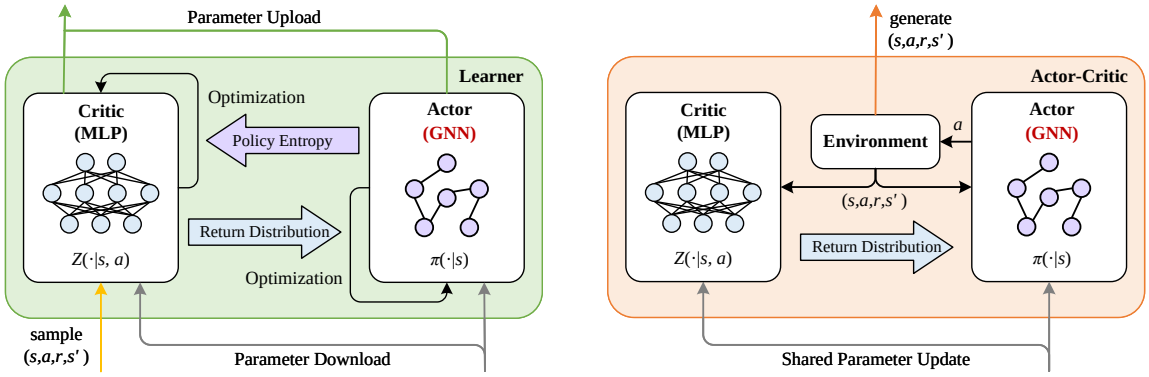

Figure 3: The structure of Learner and Actor-Critic. The return distribution and policy are approximated using two neural networks: the Actor and the Critic. DSAC initially updates the distributional value using samples collected from the buffer. Subsequently, the output of the critic network guides the update of the actor network.

$$\begin{aligned}\pi_{\text{new}} &= \arg\max_{\pi} G_{\pi} \\ &= \arg\max_{\pi} \mathrm{E}[Q^{\pi_{\text{old}}}(s,a) - \alpha \log \pi(a|s)]\end{aligned} \tag{8}$$

To prevent the overestimation of Q-values during the learning process and mitigate any potential negative impacts on policy performance, this algorithm no longer directly computes the expected value $Q^{\pi}(s,a)$ of the soft return $Z^{\pi}(s,a)$. Instead, it uses $Z^{\pi}(s,a)$. : $Z^{\pi}(Z^{\pi}(s,a)|s,a)$: $\mathrm{S} \times \mathrm{A} \rightarrow \mathrm{P}(Z^{\pi}(s,a))$. The approach is referred to as the value distribution function, which utilizes the Bellman operator as a basis for learning soft returns.

$$\begin{aligned}\mathrm{T}_{\mathrm{D}}^{\pi} Z^{\pi}(s,a) \overset{D}{=} & r + \gamma(Z^{\pi}(s_{t+1}, a_{t+1}) \\ & - \alpha \log \pi(a_{t+1}|s_{t+1})\end{aligned} \tag{9}$$

where $\overset{D}{=}$ indicates that the random variables on the left and right ends have the same probability distribution. Assuming $\mathrm{T}_{D}^{\pi} Z^{\pi}(s,a)$ follows $\mathrm{T}_{\mathrm{D}}^{\pi} Z^{\pi}(\cdot|s,a)$, the parameters are updated by minimizing the distribution distance.

$$\mathrm{Z}_{\text{new}} = \arg\min_{\mathrm{Z}} \mathop{\mathrm{E}}_{(s,a)\sim\rho_{\pi}} [d(\mathrm{T}_{\mathrm{D}}^{\pi} \mathrm{Z}_{\text{old}}(\cdot|s,a), \mathrm{Z}(\cdot|s,a))] \tag{10}$$

where d is a distance function that measures two distributions, and KL divergence is commonly used.

GNN-based Actor Strategy

Compared with MLP that does not use topological information, GNN can transfers information between nodes based on their edges. We developed a GNN-based DSAC actor, leveraging the topological information of IES, as illustrated in the Figure 3. In order to assign different importance to different neighbors, we employs GAT as actor strategy:

$$h_i^k = \alpha_{ii} \mathbf{W} h_i^{k-1} + \sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{W} h_j^{k-1} \tag{11}$$

where h_i^k represents the vector representation of node i in the k-th layer of the neural network. $\mathbf{W}$ represents the neural network parameter matrix used to linearly transform the node characteristics. $\mathcal{N}(i)$ represents the neighboring nodes of node i, and α_{ij} is the attention coefficient.

$$\alpha_{ij} = \frac{\exp(\mathrm{GELU}(\mathbf{a}^T[\mathbf{W}h_i \parallel \mathbf{W}h_j \parallel \mathbf{W}_e e_{ij}]))}{\sum_{k \in \mathcal{N}(i) \cup \{i\}} \exp(\mathrm{GELU}(\mathbf{a}^T[\mathbf{W}h_i \parallel \mathbf{W}h_k \parallel \mathbf{V}_e e_{ik}]))} \tag{12}$$

The vector $\mathbf{a}$ in the formula represents the parameter vector of the attention network. $\mathbf{W}_e$ is the parameter matrix used for linear transformation of edge information. e_{ij} represents the feature vector of the edge. $\mathrm{GELU}(\cdot)$ denotes the activation function. $\parallel$ serves as the vector connector.

In conclusion, incorporating GAT as the actor policy network for DSAC facilitates the exploration of potential node connection information in the three types of IES systems, leading to enhanced alignment between the model output and the physical constraints of IES.

Experiments

In this section, we conducted detailed training and testing on two representative IES datasets to evaluate the effectiveness of our proposed GNN-based actor Strategy.

Datasets

We evaluate our method on two datasets: Coupled System (self-made) and Large-scale System.

Coupled System is a 6-6-6 Electric-Heat-Gas IES.The coupling of diverse energy sources through CHP nodes results in the formation of a heterogeneous graph system.

Large-scale System is a modified 33-node power system and a 32-node Bali Island Thermal system (Liu et al. 2016).

Implementation Details

Due to the design constraints and the available data, when setting up the simulation environment, it is necessary to input the basic information of each node and calculate the node and edge features using the previously mentioned formulas. The input of the policy network is the current state, and the output is the actual scheduling output. The actor network consists of 4 layers with the number of neurons in each

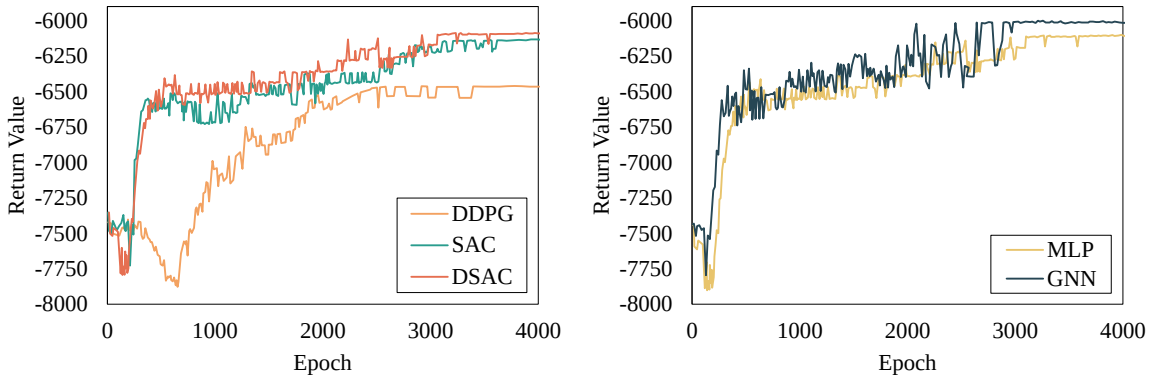

Figure 4: Return Value of Different RL algorithm (left) and DSAC based on GAT and MLP (right).

layer as follows: [128, 64, 32, 32]. The critic network consists of 5 layers with the number of neurons in each layer as follows: [128, 64, 64, 32, 32]. The experience pool size is 500000. The Adam optimizer is used to automatically adjust the learning rate within the range of $5 \times 10^{-4} \sim 5 \times 10^{-6}$.

Results and Analysis

Table 1 displays the operational average costs. After observing the performance of the model method on the two example systems, it can be concluded that DSAC outperforms other alternative reinforcement learning algorithms in terms of strategy performance, with SAC being the second-best option. On the other hand, DDPG, due to having a local optimal gradient, fails to find the global optimization solution. Firstly, by comparing the reward values of various policy network implementation frameworks, it was discovered that our proposed GAT based for implementing the DSAC policy network achieved the better performance than MLP based in two distinct example systems. These results suggest that the algorithm model based on the GNN effectively utilizes edge information, resulting in a larger exploration space, faster training speed, and avoidance of local optima.

Table 1: Reward on Coupled System and Large-scale System

Methods	**Coupled System**		**Large-scale System**	
	GAT	MLP	GAT	MLP
DSAC	**12640**	13226	**47488**	47520
SAC	13920	14268	48042	48646
DDPG	/	15582	/	49147

In the Coupled System and Large-scale System dataset, comparative experiments were conducted on DSAC, SAC and DDPG algorithms based on the MLP, with the same configuration. The return value during training are shown in left part of Figure 4. It can be observed that the DSAC algorithm exhibits advantages in terms of efficient training and excellent convergence compared to DDPG and SAC algorithms.

Under the same network settings, comparative experiments were conducted on the DSAC algorithm based on both the GAT and MLP. The rewards during the training process are shown in right part of Figure 4. The reward curve based on the GAT converges after 3000 training iterations. The reward curve based on the GAT exhibits faster convergence and higher reward values compared to the MLP.

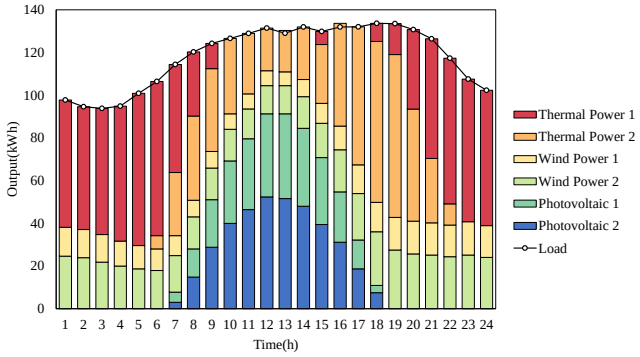

Figure 5: IES hourly output (different power) and load.

As shown in Figure 5, our method consistently aligns with the total output and load curve, even when the highest reward value is achieved. Every output adheres to the output range and ramp constraints, fulfilling the dispatch requirements of the IES. Additionally, our model incentivizes a decrease in thermal power generation output during the daytime peak of renewable energy, effectively enhancing the absorption coefficient of renewable energy without compromising the load. This achievement aligns with the goal of energy saving and environmentally friendly operation.

Conclusion

We present a DSAC reinforcement learning Optimal dispatch algorithm based on the GAT actor strategy. We demonstrate the training efficiency and effectiveness of this algorithm on two representative datasets of integrated energy system. The utilization of system topology information brings faster convergence speed and a larger exploration space compared to reinforcement learning algorithms based on the MLP. This makes it more advantageous in optimal dispatch of integrated energy systems.

References

Bronstein, M. M.; Bruna, J.; LeCun, Y.; Szlam, A.; and Vandergheynst, P. 2017. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4): 18–42.

Bruna, J.; Zaremba, W.; Szlam, A.; and LeCun, Y. 2013. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*.

Chaudry, M.; Jenkins, N.; Qadrdan, M.; and Wu, J. 2014. Combined gas and electricity network expansion planning. *Applied Energy*, 113: 1171–1187.

Duan, J.; Guan, Y.; Li, S. E.; Ren, Y.; Sun, Q.; and Cheng, B. 2021. Distributional soft actor-critic: Off-policy reinforcement learning for addressing value estimation errors. *IEEE transactions on neural networks and learning systems*, 33(11): 6584–6598.

Ge, W.; Zhang, M.; Wang, Y.; Pan, X.; and Zhou, G. 2020. Research on dynamic power flow of electric-gas coupled network based on double-time scale. *Journal of Liaoning Technical University(Natural Science)*, (2): 8.

Gori, M.; Monfardini, G.; and Scarselli, F. 2005. A new model for learning in graph domains. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 2, 729–734. IEEE.

Haarnoja, T.; Zhou, A.; Abbeel, P.; and Levine, S. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, 1861–1870. PMLR.

Huang, G.; Liu, W.; Wen, F.; Dong, C.; Zheng, y.; and Zhang, R. 2016. Collaborative planning of integrated electricity and natural gasenergy systems with power-to-gas stations. *Electric Power Construction*, 37(009): 1–13.

Li, Q.; Wang, L.; Zhang, Y.; Li, Y.; Xu, J.; and Wang, S. 2020. A review of coupling models and dynamic optimization methods for energy internet multi-energy flow. *Power System Protection and Control*, 48(19): 8.

Li, Y.; Yang, J.; Du, S. H., Xinhui; and Wang, Z. 2021. Optimal Operation of Integrated Electricityheat Energy System Considering Wind Power Consumption. *Power Capacitor Reactive Power Compensation*, 42(5): 228–235.

Liao, W.; Yu, Y.; Wang, Y.; and Chen, J. 2021. Reactive Power Optimization of Distribution Network Based on Graph Convolutional Network. *Power System Technology*, 45(6): 11.

Lillicrap, T. P.; Hunt, J. J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; and Wierstra, D. 2015. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Liu, J.; Gao, F.; and Luo, X. 2019. Survey of deep reinforcement learning based on value function and policy gradient. *Chinese Journal of Computers*.

Liu, Q.; Zhai, J.; Zhang, Z.; Zhong, S.; Zhou, Q.; Zhang, P.; and Xu, J. 2018. A survey on deep reinforcement learning. *Chinese Journal of Computers*, 41(1): 27.

Liu, X.; Wu, J.; Jenkins, N.; and Bagdanavicius, A. 2016. Combined analysis of electricity and heat networks. *Applied Energy*, 162: 1238–1250.

Liu, Y.; Dong, L.; Wang, C.; Li, M.; Qiao, J.; and Wang, X. 2021. Coordinated Optimization of Integrated Electricity-Heat Energy System Based on Soft Actor-Critic. *Smart Grid*, 11(2): 11.

Mazyavkina, N.; Sviridov, S.; Ivanov, S.; and Burnaev, E. 2021. Reinforcement learning for combinatorial optimization: A survey. *Computers Operations Research*, 134(1): 105400.

Micheli, A. 2009. Neural network for graphs: A contextual constructive approach. *IEEE Transactions on Neural Networks*, 20(3): 498–511.

Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; and Riedmiller, M. 2013. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Qiao, J.; Wang, X.; Zhang, Q.; Zhang, D.; and Pu, T. 2021. Optimal Dispatch of Integrated Electricity-gas System With Soft Actor-critic Deep Reinforcement Learning. *Proceedings of the CSEE*, 41(3): 14.

Qiu, J.; Yang, H.; Dong, Z. Y.; Zhao, J. H.; Meng, K.; Luo, F. J.; and Wong, K. P. 2015. A linear programming approach to expansion co-planning in gas and electricity markets. *IEEE Transactions on Power Systems*, 31(5): 3594–3606.

Salimi, M.; Ghasemi, H.; Adelpour, M.; and Vaez-ZAdeh, S. 2015. Optimal planning of energy hubs in interconnected energy systems: a case study for natural gas and electricity. *IET Generation, Transmission & Distribution*, 9(8): 695–707.

Tang, M.; Luo, Y.; Hu, B.; Zhu, W.; Li, T.; and Dou, W. 2020. A review of the dispatch model of a combined heat and power system. *Power System Protection and Control*, 48(23): 15.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; and Yu, P. S. 2021. A Comprehensive Survey on Graph Neural Networks. *IEEE transactions on neural networks and learning systems*, (1): 32.

Yang, H.; Li, M.; Jiang, Z.; Liu, X.; and Guo, Y. 2020. Optimal operation of regional integrated energy system considering demand side electricity heat and natural-gas loads response. *Power System Protection and Control*, 48(10): 8.

Yang, Z.; Huang, s.; and Chen, Y. 2021. Research on cooperative optimal operation of multi-park integrated energy system based on multi agent reinforcement learning. *Advanced Technology of Electrical Engineering and Energy*, 40(8): 10.

Zhou, H.; Liu, G.; and Liu, C. 2014. Study on the Energy Internet Technology Framework. *Electric Power*, 47(11).