Design Studies and Intelligence Engineering
L.C. Jain et al. (Eds.)

doi:10.3233/FAIA250338

Deep Learning-Based Optimization Study of Virtual Power Plant Resource Regulation

Qilin CAO[a1],Shuying ZHANG[a], Zhanqi GU[a],Huaming ZHOU[a],Leting YU[a]
[a]*State Grid Shanghai Electric Power Company,Shanghai 200000,China*

Abstract. In order to solve the problem that it is difficult to protect privacy in the resource optimization and control of virtual power plants, the research on resource optimization and control of virtual power plants based on deep learning was proposed. In this paper, a subarea distributed optimal regulation framework based on multi-agent deep reinforcement learning is proposed, and a day ahead optimal regulation model is constructed with the goal of minimizing the daily operation cost, and various operation constraints are considered. The near end strategy optimization algorithm is used to train the model offline, and the trained model is used to make online optimal control decisions. The improved IEEE33 node example is used for simulation verification. The experimental results show that the control cost of the proposed method is 37.53% less than that of the particle swarm optimization algorithm, while it is basically consistent with the centralized mathematical planning method, only 1.66% more than that of the centralized mathematical planning method. In terms of decision-making time, through 40 min offline training, the online decision-making time of multi-agent in-depth reinforcement learning is only about 0.1840s, which is significantly higher than the online decision-making time of centralized mathematical planning method. Conclusion: The centralized optimization effect can be achieved only through the collection of information in this region and the sharing of hidden layer feature information of neural networks between adjacent regions, which solves the problem that the centralized system is difficult to obtain the private data of each autonomous region.

Keyword.virtual power plant resources; distributed optimal scheduling; multi-intelligentsia; deep reinforcement learning

1. Introduction

In recent years, many countries have been exploring the energy development model combining Internet technology and renewable energy technology. With smart grid as the resource allocation center, they have built a multi energy coupling system with horizontal multi-source complementation and vertical source network load storage coordination. Energy Internet came into being [1]. Under the framework of energy Internet, the energy supply and demand system is a "source network load storage" complex, where multiple energy sources such as electricity, heat, cold and gas are coupled and interconnected from the demand side to the supply side [2]. In this context, FRs not only include distributed

[1] Corresponding Author.Qilin CAO,E-mail:yqcql2003@126.com.

generation resources (DG) such as wind, light and biomass, and distribution side and demand side resources such as energy storage, electric vehicles and controllable loads, but also can accommodate a variety of energy coupling equipment, including cold, heat and power cogeneration units, gas boilers, gas absorption chillers, electric boilers and electric chillers [3].

Virtual power plant (VPP), as an effective means of aggregating FRs, can realize energy gathering, energy storage, energy supply and energy consumption without changing the grid connection mode and geographical location of FRs by virtue of advanced communication, measurement, control and other technologies, effectively connect FRs with the power system, realize resource integration and distribution, and gradually participate in the operation of distributed energy in the wholesale power market as a aggregating entity, It is an important way for smart grid to realize interaction and intelligence on the energy supply and demand side [4]. Compared with the horizontal single energy composition of traditional power plants, virtual power plants incorporate a large number of renewable energy, energy storage, controllable load and other distributed energy, break the monopoly from the system, and build a power plant architecture combining horizontal multiple complementation and vertical source network load storage coordination [5]. In addition, the respective advantages of different FRs can be used to provide multiple services for the power grid, such as energy balance, reactive power/voltage support, rotating reserve, frequency regulation and congestion management, showing a certain economic value [6].

Through the comprehensive review of the virtual power plant scale flexible resources and the systematic study of its grid controllability, combined with the resource grid access system, a virtual power plant scale flexible resource aggregation and control framework with comprehensive resource coverage and clear aggregation and control paths has been formed [7].

2. Literature review

In recent years, with the growing popularity of distributed energy, its significant volatility and uncertainty pose new challenges to the safe and stable operation and economy of power systems [8]. The distributed resources on the national demand side are very rich. Although distributed resources such as air conditioning load, energy storage equipment and data center can use their flexible regulation potential to provide regulation services for the power grid, their regulation potential cannot be effectively used because their monomer volume is small and their locations are scattered, and they cannot accept the direct regulation of the power grid [9]. At present, more and more attention has been paid to the aggregation of demand side distributed resources to achieve centralized regulation [10]. The virtual power plant can aggregate large-scale distributed resources and conduct centralized regulation, so as to realize the rational optimal allocation and utilization of resources [11].

The centralized optimization method collects the data of the whole network through the control center for optimization and control calculation, and sends the control instructions to each controlled unit [12]. et al. established a model with microgrid and distribution access point active as the decision variables, and the objectives of minimizing distribution network loss and optimizing voltage quality, and the solution method is a non-dominated genetic algorithm with elite strategy [13]. et al. solved the proposed source-network-load-storage two-layer optimization model with a multi-

objective particle swarm algorithm and a two-quantum differential evolutionary algorithm, respectively [14]. However, the centralized optimization method requires each microgrid or distribution network autonomous region to share internal information completely, distributed resources belong to different interests, and the internal equipment operation data and sensitive user loads in each region are private data, so the centralized control system is unable to collect detailed information of the controlled objects or control each region directly [15].

In view of the difficulty of privacy protection in centralized optimal regulation of distributed generation, this paper proposes a research on resource regulation and optimization of virtual power plants based on deep learning. Firstly, the basic principle of MADRL method based on communication neural network (CommNet) is described. Then, a distributed optimal regulation framework based on MADRL is proposed, and a day ahead optimal regulation model is constructed with the goal of minimizing the daily operation cost of active distribution network. Then, the model is trained using the proximal policy optimization (PPO) algorithm, and the trained model is used to make online optimal control decisions. Finally, an example shows that each autonomous region can calculate its own approximate global optimal solution under the condition of only using local communication. This method can adapt to the uncertainty of distributed generation output and load, and give real-time optimal control results according to the random changes of source load.

3. Methodology

3.1 Principles and Methods of Deep Reinforcement Learning for Multi-Intelligents

MADRL takes the multi-agent Markov decision process (MAMDP) as the basic framework, and MAMDP can be expressed as a tuple$\langle N, \{S^n\}, \{A^n\}, P, R, \gamma\rangle$. Among other things$N$is the number of intelligences.$S^n$for intelligent bodies$n$The set of states of $s_t^n \in S^n$ Indicates that the intelligent body is in, the$t$time period, the states of all the intelligences are united together to form the joint state vector $S = S^1 \times S^2 \times ... \times S^N$And there is $s_t \in S; A^n$ It's an intelligent body$n$The set of actions, the$a_t^n \in A^n$Indicates that the intelligent body is in, the$t$The actions selected in the time period,the actions of all the intelligences are combined together to form a joint action vector $\boldsymbol{A} = A^1 \times A^2 \times ... \times A^N$ And there is $\boldsymbol{a}_t \in \boldsymbol{A}; P: \boldsymbol{S} \times \boldsymbol{A} \times \boldsymbol{S} \rightarrow [0,1]$ is the state transfer probability,denoted at, the$s_t$under which the action is executeda_tAfter the state of the environment is transferred to that$s_{t+1}$the probability of; the$R$is the reward function,denoted at, the$s_t$under which the action is executeda_tAfter the environment gives rewards, and there are$r_t \in R; \gamma$for the discount factor. During the time period, the$t$The intelligences act on the environment by executing actions based on the observed states and obtaining rewards from the environmentr_t, and to change the state of the environment to$S_{t+1}$and so interacting with the environmental loop, using the resulting data to modify the joint action strategy, i.e., the mapping relationship between joint actions and joint states $\pi: S \rightarrow A \Rightarrow a \sim \pi(a \mid s)$, to maximize cumulative returns [16].

CommNet is a multi-agent deep reinforcement learning neural network architecture, as shown in Figure 1. Each agent has a neural network with the same structure to make decisions$t$For NonAn intelligence whose network input is its own observation state to protect privacy s_t^n, with the output being the decision action a_t^n. In the encoding layer,the

encoding function that $h_{t,0}^n = r(s_t^n)$ The input information is transformed into hidden state information into the communication layer. The communication layer is the key to realize the collaboration among the intelligences, and each intelligence transforms the hidden layer state information, the $h_{t,m-1}(m$ On behalf of No m communications)into the communication layer network f_m, and to the outputs of the communication layer networks of neighboring intelligences $h_{t,m}$ Doing mean pooling, the results obtained and $h_{t,m}$ as the input to the next layer of neural network for each neighboring intelligence [17]. The iterative relationship between the inputs and outputs of each layer of the communication layer of each intelligent is shown in equation (1)(2) as follows.

$$h_{t,m}^n = f_m\left(h_{t,m-1}^n, c_{t,m-1}^n\right) = \sigma\left(H_m h_{t,m-1}^n + C_m c_{t,m-1}^n\right) \tag{1}$$

$$c_{t,m}^n = \frac{1}{|\mathcal{N}(n)|}\sum_{n' \in \mathcal{N}(n)} h_{t,m}^{n'} \tag{2}$$

Where. H_m and C_m Indicates that the first m Layer communication layer network parameters to be updated. σ denotes the nonlinear activation function; the $\mathcal{N}(n)$ Representation and Intelligent Bodies$n$The set of neighboring intelligences. The last layer of the network is the decoding layer, which transforms the hidden layer state information into actual actions.

The neural network architecture can be trained using different depth reinforcement learning algorithms. The strategy gradient algorithm based on the Actor-Critic framework can obtain good decision results for continuous action space problem, and distributed generation optimization and regulation is just such a problem [18]. PPO algorithm is a policy gradient algorithm based on Actor Critic framework, which solves the problem of low data sampling efficiency of traditional policy gradient algorithm. Therefore, this paper plans to use PPO algorithm to train the above CommNet network.

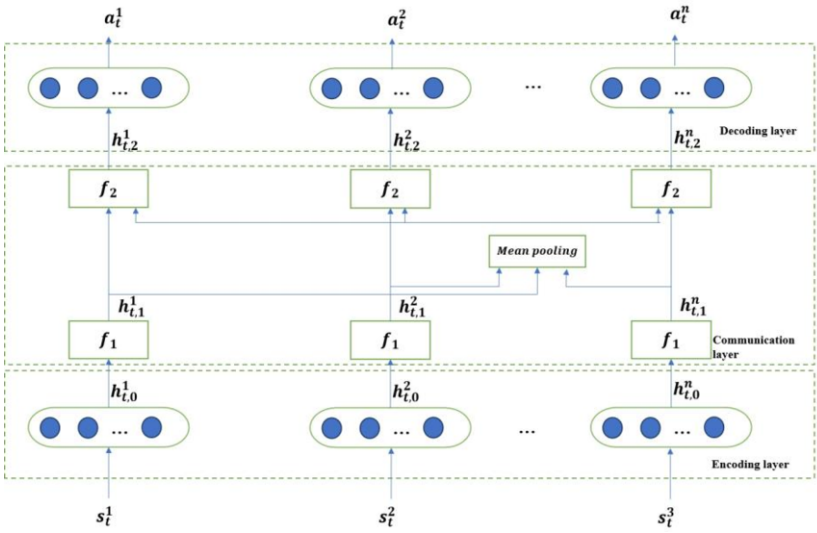

Figure1. CommNet Network Structure

PPO algorithm uses strategy (Actor) neural network to approximate strategy $\pi(a \mid s)$, the parameters are denoted asθ_{a}; Use the Critical neural network to approximate the value function $V_{\pi}(s)$, which evaluates the actions selected by the policy network and thus guides the policy network update, is parameterized as θ_{c}。 $V_{\pi}(s)$ denotes the states$s$The expected value of the cumulative return can be obtained under equation (3).

$$V_{\pi}(s) = E_{\pi}(\sum_{k=0}^{\infty} \gamma^{k} r_{t+k} \mid s_t = s) \tag{3}$$

Sampling data obtained from the interaction of the unupdated strategy network with the environment (s_t, a_t, r_t, s_{t+1}) which are stored in the experience pool and can be used multiple times for new strategies $\pi(a \mid s)$ The data were updated in order to improve the efficiency of data sampling.

3.2 Distributed Optimization Regulatory Framework

A distributed optimal regulation framework is proposed based on a multi-intelligence deep reinforcement learning approach. The active distribution network with high penetration of distributed power is divided into several autonomous regions [19]. Each autonomous region sets up a regional intelligent body, which is responsible for collecting renewable energy generation forecasts, load forecasts, and state information of each device in the region, and issuing regulation commands to the controllable devices in the region. Neighboring regional intelligences can interact with each other [20].

In this paper, the PPO algorithm is used to train the agent neural network in each region, and the network parameters are obtained. Section$t$Regional intelligences$n$Based on the state of the regionss_t^n, using strategy networks to output regulatory decisionsa_t^n:: Because autonomous regions operate in concert, regional intelligences receive the same global rewards based on the regulatory decisions of all regional intelligencesr_t^n, and move on to the next regulation period for decision making. Taking the experience gained from sampling each regulation period, the(s_t, a_t, r_t, s_{t+1})stored in the experience pool to maximize each regulation cycle$T$The cumulative global reward over time periods is targeted to update the network parameters.

Each autonomous region can achieve autonomy in its own region, and can also operate cooperatively. On the one hand, each regional agent contains a strategic network and a value network, which can realize the local data processing of distribution networks with high penetration of distributed generation; On the other hand, the strategy network and value network are connected by CommNet architecture, so as to interact the hidden layer feature information of neural network among agents in adjacent regions, and realize distributed collaborative optimization of multiple autonomous regions. The centralized system is difficult to collect the detailed information of each distributed resource because it involves the owner's privacy. Under this architecture, the agents in each region can realize the cooperative solution of the problem based on the information collected locally and the communication between the agents in adjacent regions, avoiding the transmission of a large amount of private data in the regulation process.

3.3 Algorithmic flow

The PPO algorithm is used for offline training of the above optimization control model. The overall process of the algorithm is shown in Figure 2, and the specific steps are as follows.

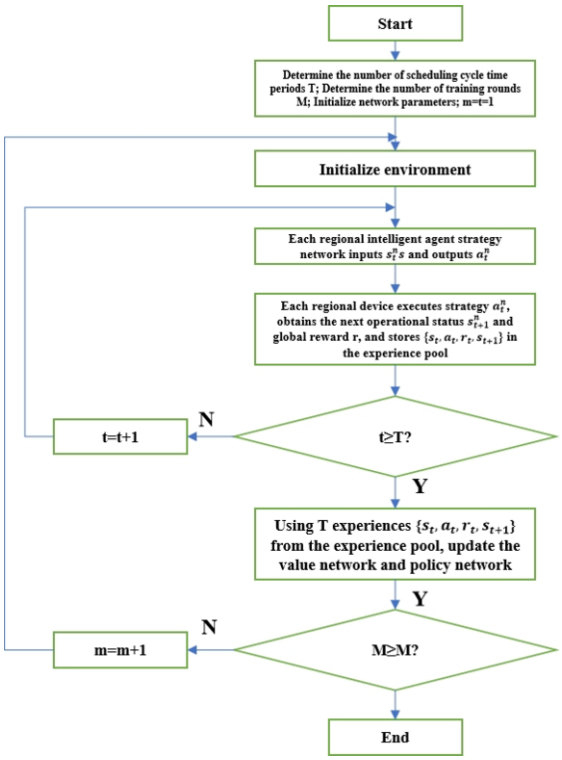

Figure 2. MADRL algorithm flow

Step 1: Determine the total number of time slots in each regulatory cycle$T$and the number of rounds of neural network training for regional intelligences$M$, and randomly initialize the parameters of the policy network of the regional intelligences$\theta_{\mathrm{a}}$and value network parametersθ_{c}。

Step 2: initialization of the environment, to 00:00 every day as the initial regulation moment, the state of the equipment before the regulation of the moment is initialized, including the micro-gas turbine output in equation (4)$P_0^{n,mt}$and electrical energy storage capacity$s_0^{n,b}$。

Step 3: Each regional intelligent body interacts with the environment and collects information about the state of the region as shown in equation (4)s_t^nAs a network input, the action information shown in Eq. (6) is outputa_t^n, as a micro-gas turbine and electric energy storage regulation command, and the global reward shown in equation (22) is calculated at the end of this time period of regulationr_t;Combining the sampling experience of each time period(s_t, a_t, r_t, s_{t+1})Stored in an experience pool for network parameter updates.

Step 4: Regional intelligences training. At the end of a moderation cycle sampling, using the experience pool, the$T$In this study, the gradient descent method is used to update the strategy network and value network of each intelligent body, and the update goal is to maximize the global reward accumulated in one regulation cycle$\sum_{t=1}^{T}\gamma^{t-1}r_t$, the two types of network learning rates are denoted, respectively, asl_{a}andI_{c}。

Step 5: Determine whether the set maximum number of training rounds has been reached$M$,if satisfied then end the training; if not satisfied then return to step 2 for the next round of network parameter update.

3.4 Simulation analysis

In order to verify the effectiveness of the proposed distributed optimal regulation method based on multi-agent deep reinforcement learning, the improved IEEE 33 node active distribution network is used as an example for simulation research, and the voltage level is 12.66kV.

The test system was divided into 3 autonomous regions. The PVs are distributed in nodes, respectively $\{15,19,29\}$; micro gas turbines are distributed at nodes, respectively $\{13,24,32\}$; electrical energy storage is distributed in nodes, respectively$\{10,21,27\}$。 The total PV and total load power of each region are taken from the data of every 15min for two months from July 1, 2020 to August 31, 2020 in three different regions of the Belgian power grid. The regulation period is 24 hours, with 15 minutes as one regulation period, and one day is divided into 96 periods. Time of use electricity price is adopted in this paper, in which the peak period is$08:00-19:00$"The usual paragraph is$06:00-8:00,19:00—23:00$The valley hours are$23:00-06:00$。

The state observation of each regional intelligence is represented as a 5-dimensional array vector, and the action is represented as a 4-dimensional array vector. The coding layer and communication layer of each intelligent's strategy network and value network have the same structure, in which the coding layer has one layer containing 128 neurons, and the communication layer has two hidden layers, and the number of neurons contained in each layer is also 128. For the decoding layer of the strategy network, there are three hidden layers, and the number of neurons in each layer is, respectively128、64、4; the number of neurons in the 3 hidden layers of the decoding layer of the value network are, respectively, the number of128、64、1。 All hidden layers use the Tanh activation function. The learning rates of the policy network and the value network are respectively taken as$l_a = 0.01, I_c = 0.001$, the discount factor is taken as$\gamma = 1$The In the normal distribution obeyed by the prediction bias, the$\mu = 0$For photovoltaic forecast bias.εTake that10%, for load forecast deviations.εTake that3%。

The program in this paper is written based on the tensorflow1.14 framework. The computing hardware condition is Core i3-9100F CPU, 3.60GHz, 8G memory, and the number of algorithm iterations is 32000.

4. Results and discussion

In order to verify the effectiveness of the multi-intelligence deep reinforcement learning method in this paper, the following three optimization methods are used for comparative analysis.

Method 1:The multi-intelligence deep reinforcement learning method proposed in this paper.

Method 2:Select particle swarm optimization algorithm for comparative analysis.

Method 3: Mathematic planning method, the interior point optimizer (IPOPT) is selected to solve the optimal regulation problem of active distribution network.

The comparison results are shown in Table 1. It can be seen that compared with the centralized optimization method, the method proposed in this paper has the following advantages: 1) In terms of optimization cost, the control cost of the method proposed in this paper is 37.53% less than that of the particle swarm optimization algorithm, while it is basically consistent with the centralized mathematical planning method, which is only 1.66% more than that of the centralized mathematical planning method. 2) In terms of decision-making time, through 40 min offline training, the online decision-making time of multi-agent in-depth reinforcement learning is only about 0.1840s, which is significantly higher than the online decision-making time of centralized mathematical planning method. In the proposed method, the optimization calculation time is assumed by the network training phase, and the online decision-making phase is directly based on the trained network from input to output. 3) In addition, the method proposed in this paper has a distributed feature. During the training process, agents in each region do not need to share photovoltaic, load forecasting data, equipment parameters and other information, which can protect the privacy of each region.

Table 1. Comparison of methods

Methods	Cost/$	Decision time/s
The methodology of this paper	9821.81	0.184
particle swarm algorithms	15721.846	1213.084
IPOPT	9661.43	10.138

5. Conclusion

In this paper, we propose a deep learning-based optimization study of virtual power plant resource regulation. Aiming at the problem of privacy protection that is difficult to be solved in the optimization and regulation of virtual power plant resources, this paper proposes a distributed optimization and regulation method based on multi-intelligence deep reinforcement learning, which has the following advantages: centralized optimization can be achieved only through the collection of information in the region and sharing of the neural network hidden layer feature information between neighboring regions. It solves the problem that it is difficult for the centralized system to obtain the privacy data of each autonomous region. It can adapt itself to the uncertainty of distributed power supply and load, and give optimization and control results in real time according to the stochastic change of source and load without relying on the accurate prediction of source and load. It realizes the cooperative operation of multiple autonomous regions of the active distribution network, which greatly improves the overall economy of the system compared with the independent operation of each region.

References

[1] Jia, D. , Shen, Z. , & Li, X. L. X. . (2023). Bi-level scheduling model for a novel virtual power plant incorporating waste incineration power plant and electric waste truck considering waste transportation strategy. Energy conversion & management, 298(Dec.), 1-18.

[2](2024). Distributed energy resource integration for carbon neutral power systems: market-based approaches to ancillary services and microgrid operation. IEEJ Transactions on Electrical and Electronic Engineering, 19(5), 598-607.
[3] Sharma, H. , & Mishra, S. . (2022). Optimization of solar grid?based virtual power plant using distributed energy resources customer adoption model: a case study of indian power sector. Arabian journal for science and engineering, 47(3), 2943-2963.
[4] Laezman, R. . (2023). Virtual power plants charging up on renewables. Electrical contractor, 88(10), 12-12.
[5] Pirouzi, S. . (2023). Network-constrained unit commitment-based virtual power plant model in the day-ahead market according to energy management strategy. IET generation, transmission & distribution, 17(22), 4958-4974.
[6] Ajmera, K. , & Tewari, T. K. . (2023). Energy-efficient virtual machine scheduling in iaas cloud environment using energy-aware green-particle swarm optimization. International Journal of Information Technology, 15(4), 1927-1935.
[7] Wang, S. , Wu, W. , Chen, Q. , Yu, J. , & Wang, P. . (2024). Stochastic flexibility evaluation for virtual power plants by aggregating distributed energy resources. CSEE Journal of Power and Energy Systems, 10(3), 988-999.
[8] Tamilselvan, K. , Kaliappan, L. , & Kandasamy, P. . (2023). An optimal selection and placement of distributed energy resources using hybrid genetic local binary knowledge optimization. Journal of circuits, systems and computers, 32(18), 1-21.
[9] Gao, H. , Zhang, F. , Xiang, Y. , Ye, S. , Liu, X. , & Liu, J. . (2023). Bounded rationality based multi-vpp trading in local energy markets: a dynamic game approach with different trading targets. CSEE Journal of Power and Energy Systems, 9(1), 221-234.
[10] Akkas, O. P. , & Cam, E. . (2023). Risk-based optimal bidding and operational scheduling of a virtual power plant considering battery degradation cost and emission. advances in electrical and computer engineering, 23(2), 19-28.
[11] Ghanuni, A. , Sharifi, R. , & Farahani, H. F. . (2023). A risk-based multi-objective energy scheduling and bidding strategy for a technical virtual power plant. Electric Power Systems Research, 220(Jul.), 1-12.
[12] Yang, Z. , Li, K. , & Chen, J. . (2024). Robust scheduling of virtual power plant with power-to-hydrogen considering a flexible carbon emission mechanism. Electric Power Systems Research, 226(Jan.), 1-13.
[13] Song, J. , Yang, Y. , & Xu, Q. . (2023). Two-stage robust optimal scheduling method for virtual power plants considering the controllability of electric vehicles. Electric Power Systems Research, 225(Dec.), 1-23.
[14] Zhou, D. , & Zhu, Z. . (2023). Urban integrated energy system stochastic robust optimization scheduling under multiple uncertainties. energy reports, 9(Suppl.7), 1357-1366.
[15] Kang, H. , Jung, S. , & Hong, J. T. . (2024). Reinforcement learning-based optimal scheduling model of battery energy storage system at the building level. Renewable & sustainable energy reviews, 190(Feb. Pt.A), 1-16.
[16] Zhang, Y. , Liu, F. , & Wang, W. F. S. . (2022). Robust scheduling of virtual power plant under exogenous and endogenous uncertainties. IEEE Transactions on Power Systems: A Publication of the Power Engineering Society, 37(2), 1311-1325.
[17] Wan, X. , Wen, X. , & Gong, S. S. Q. . (2023). The low-carbon economic operation strategy of virtual power plant under different electricity-gas-heat-carbon multi-market synergy scenarios. Journal of computational methods in sciences and engineering, 23(4), 2237-2254.
[18] Wan, X. , Wen, X. , & Tang, B. S. Q. . (2023). Trading strategy for virtual power plant clusters based on a multi-subject game model. Journal of computational methods in sciences and engineering, 23(5), 2261-2274.
[19] Ullah, Z. , & Baseer, M. . (2022). Operational planning and design of market-based virtual power plant with high penetration of renewable energy sources. International Journal of Renewable Energy Development, 11(3), 620-629.
[20] Mangalampalli, S. , Swain, S. K. , & Mangalampalli, V. K. . (2022). Multi objective task scheduling in cloud computing using cat swarm optimization algorithm. Arabian journal for science and engineering, 47(2), 1821-1830.