Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 08, August-2021

Profit-based Units Scheduling of a GENCO in Pool Market using Deep Reinforcement Learning

Emmanuel G. S
M.Tech Student: Department of Electrical and Electronic Engineering, JNTUA-CEA
Andhra Pradesh-India

Prof. P. Sujatha
Professor: Department of Electrical and Electronic Engineering, JNTUA-CEA
Andhra Pradesh-India

Dr. P. Bharath
Assistant Professor:
Department of Electrical and Electronic Engineering,
JNTUA-CEA
Andhra Pradesh-India

Abstract: **The primary objective of power generation unit scheduling for a GENCO operating in restructured power system, is to maximize accumulated profit over the entire period of operation. When operating in the pool market, GENCO's demand is the spot market allocated energy. Hence, prior to units scheduling, the GENCO has to forecast how the market will be as far as the market clearing price and the spot market allocation for each hour of the day is concerned. By using these two market signals, the company can optimally schedule its generation to maximize profit. However, this paper aims at exploring capability of Deep Reinforcement Learning (DRL) established by using Deep Deterministic Policy Gradient (DDPG) algorithm to optimally schedule generating units in order to boost GENCO's financial benefit in deregulated electricity market environment. Simulations were carried out for a GENCO with six generating units each with different operation cost curve and different generating capacity, the resulted output reveal that the proposed method can be applied to solve profit-based generation units scheduling problem (PBUS).**

Keywords: DDPG, DRL, Market Clearing Price (MCP), Profit based unit scheduling (PBUS), Power system deregulation.

Abbreviations

$C(P_{it})$	cost of generating P amount of power at hour t by i^{th} generating unit
i	i^{th} generating unit
N	total number of generating units
PF_t	total profit at hour t
Pen_t	penalty at hour t
$pMCP$	predicted market clearing price
P_{it}	power generated by i^{th} unit at hour t
pMA_t	predicted Market allocation
Rev_{it}	revenue at time t of the i^{th} generating unit
R_t	overall reward at hour t
SUC_{it}	Star-tup cost of unit i^{th} at time t
$P_{i,max}$	Generator maximum generated power
$P_{i,min}$	Generator minimum generated power

I. INTRODUCTION

The deregulation process in energy sector is one of the most important transition for modern electricity industry. This transition enhance the competition in the electricity market the power prices are likely to descend which favours the electric power consumers [1],[2],[3]. With such idea in mind, there is a need to optimally schedule the generation units in a manner that will generate more profit [4]. This is due to the fact that, this type of market is based on competition which affects the electricity energy price. In contrast from vertical integrated power system, where utilities had obligation to meet demand and reserve, in deregulated power system the main objective of GENCO is to maximize its profit [5],[6],[7]. That is, GENCO has to schedule its generation pattern that will maximize the total profit. On the other hand, the responsibility of Independent System Operator (ISO) is to satisfy the system power demand in order to balance between generation and load. The ISO neither owns nor operates any generating unit but receives bids from different GENCOs and it decides energy demand among the GENCOs based on a cheapest first method [8].

UC problem has been solved by several methods each with its advantage(s) and disadvantage(s), [6],[9] explained the priority list method, dynamic programming [10],[11], Lagrangian relaxation, Genetic Algorithm [12], Grey wolf optimization [13], Particle Swarm optimization [14], Tabu Search method, Fuzzy logic algorithm [15] and Evolutionary algorithm [16]. However, a few reactions have been routed to these strategies as they are iterative require an initialization step. That can cause the convergence property for the pursuit interaction into ideal local optimal solution. Also, they may neglect to tackle the powerful case including above limitations. Market clearing price and the load forecasts plays an important part in strategizing optimal bidding in a day ahead market[5], [6].

The reinforcement learning technique has been used to solve complex problems and high dimensional problems in control systems [19], delivery route problem [20] and robotics [21]. the aim of this study is to introduce the use of deep reinforcement learning to solve optimally scheduling of generating units scheduling in deregulated power system environment in order to maximize GENCO's profit. By analysing the operation predicted data (the market clearing prices and market allocations), a data-driven Profit based

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 08, August-2021

unit scheduling (PBUS) model is established. By means of DDPG algorithm, the established model is trained to maximize the GENCO's profit, finally the model is tested to show the effectiveness and accuracy of the proposed method.

II. PROBLEM FORMULATION

The objective of the PBUS problem is to formulate a scheduling pattern that will maximize the expected profit for the entire operation period. Therefore, the objective function is expressed as the difference of revenue generated and cost spent [22],[23]. The optimization problem for PBUS can be formulated mathematically by the following Equations;

Objective function

$$\max PF_T = Rev_T - FC_T \tag{1}$$

Where;

$$Total\ Revenue\ RV_T = \sum_{i=1}^{n}\sum_{t=1}^{T}(P_{it} \times pMCP) \tag{2}$$

$$Total\ Fuel\ Cost\ FC_T = \sum_{i=1}^{n}\sum_{t=1}^{T}(a_i + b_i P_{it} + c_i P_{it}^2 + SUC_{it}) \tag{3}$$

Constraints;

$$\left(\sum_{i=1}^{n} P_{it} U_{it}\right) \le pMA_t\ ; \forall t \in T \tag{4}$$

$$P_{i,min} \le P_i \le P_{i,max}\ ;\ \forall i \tag{5}$$

III. THE PROPOSED METHOD

A. Reinforcement Learning

Reinforcement learning is a class of machine learning, that is based on trial-and-error, that is concerned with sequential decision making [24]. An RL agent exists in an environment. Within the environment it can act, and it can make observations of its state and receive rewards. These two discrete steps, action and observation, are repeated indefinitely with the agent's goal being to make decisions so as to maximize its long-term reward.

B. Deep reinforcement learning

DRL utilize deep neural net as function approximator, which are especially valuable in reinforcement learning in the case that observations and/or actions dimension are so high that one can't even think about being totally known [25], [26]. In the deep reinforcement learning, deep neural network is utilized to implement either a value function, or a policy function i.e, networks can figure out how to get values for the given states, or getting values from actions and observations sets. Instead of using the technique of Q-table which would be very expensive method one can train a neural network from the given dataset of states or actions examine how significant those are comparative with our target in reinforcement learning [27].

Like every neural network, coefficients are used to estimate the function relating inputs to outputs, and their learning comprises to tracking down the correct coefficients, or weights, by iteratively changing those weights along gradients that guarantee less error. In reinforcement learning, convolutional organizations can be utilized to perceive a specialist's state when the input is visual images. Figure 1 is an architecture used to design both actor and target actor networks. The network model constituted of; feature input layer, four fully connected layers (Which are all feed forward neural networks) and four activation functions.

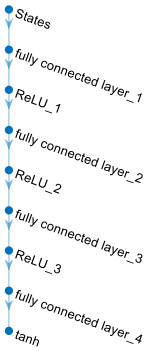

Figure 1. Actor and target actor structure

Figure 2 shows the architecture of the critic network of the DRL used, the same is applied for target critic network. It can be seen that the critic network receives both observations and actions from the actor and output the Q values.

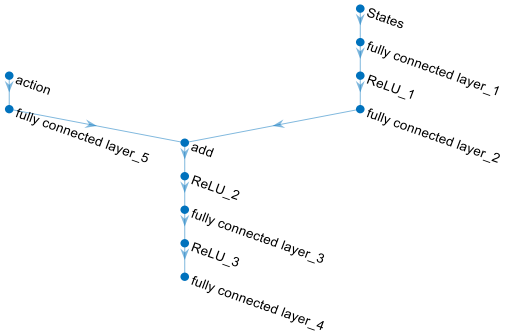

Figure 2. The critic and target critic structure

C. Designing of the state space, action space and reward function

Reward is defined as profit plus penalties, the agent is penalized when the sum of power generated is more than the predicted market allocation (pMA_t).

First reward to be considered is Profit at each time step which is calculated as;

$$PF_t = \sum_{i=1}^{N}(Rev_{it} - C(P_{it})) \tag{6}$$

Whenever the agent violates the constraints defined it should be penalized as per equation (7)

$$Pen_t = \begin{Bmatrix} -k\ if\ (\sum_{i=1}^{i=N} P_{it}) > pMA_t \\ 0\ otherwise \end{Bmatrix} \tag{7}$$

Net reward at each time step is defined as sum of penalty and profit as per equation (8)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 08, August-2021

$$R_t = PF_t + Pen_t \quad (8)$$

Defining Agent's states

States = $\{\hat{P}_{1t,}\ \hat{P}_{2t,} \ldots\ \hat{P}_{it,} \ldots\ \hat{P}_{Nt}\}$

Where

$$\hat{P}_{it} = Pit - \left(\frac{pMCP_t - bi}{2c_i}\right) \quad (9)$$

Defining Agent's action

Actions = $\{\ P_{1t}, \ldots\ Pit,\ \ldots\ P_{Nt}\ \}$

D. The DDPG Algorithm

The algorithm uses a total of four neural networks. The first network is called the actor, $\pi(s|\theta^{\pi})$, where θ^{π} denotes the network parameters. The actor part of the DDPG agent is classified as a policy search method.

The second network is called the critic, $Q(s,a|\theta^{Q})$, where θ^{Q} denotes the network parameters. The critic part of the DDPG agent is classified as a value function method [28],[29].

DDPG uses target network idea to implement further two neural networks, one for each of the actor and critic networks. The network parameters for the actor and critic target networks are denoted as $\theta^{\pi'}$ and $\theta^{Q'}$ respectively. DDPG also makes use DQN's experience replay buffer to store experience which is randomly sampled from during training [30].

The loss function for the critic network is similar to the DQN loss function except that actions are selected by the actor network [31]. Using the standard Q-learning update and the mean square error, the critic loss function is expressed as:

$$L_i(\theta_i^{Q}) = E_{(s,a,r,s')\sim U(D)}[(r + \gamma Q(s', \pi(s'|\theta_i^{\pi'})|\theta_i^{Q'}) - Q(s,a|\theta_i^{Q}))^2] \quad (10)$$

The actor network is updated using the deterministic policy gradient theorem [31]. The gradient update is given by:

$$\nabla_{\theta^{\pi}} J(\theta^{\pi}) = E[\nabla_a Q(s,a|\theta^{Q})|_{s,a=\pi(s|\theta^{\pi})} \nabla_{\theta^{\pi}} \pi(s|\theta^{\pi})|s] \quad (11)$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 08, August-2021

Table 1. Actor and critic parameter settings

	Feature Input Layer	Fully connected layer 1	Fully connected layer 2	Fully connected layer 3 and 5	Fully connected layer 4
Input size	6	6	100	100	100
Output size	6	100	100	100	6 for actor network 1 for critic network
Number of hidden layers	n/a	32	64	64	32
Weight learning rate factor	n/a	1	1	1	1
Regularization factor for weights	n/a	1	1	1	1
Bias learning rate factor	n/a	1	1	1	1
Regularization factor for biases	n/a	0	0	0	0
Weight initializer	n/a	Glorot	Glorot	Glorot	Glorot
Bias initializer	n/a	Zeros	Zeros	Zeros	Zeros
Activation function	n/a	ReLU	ReLU	ReLU	tanh

Equations (10) and (11) are used with gradient descent and the backpropagation algorithms to update actor and critic network weights during training. The algorithm flowchart is

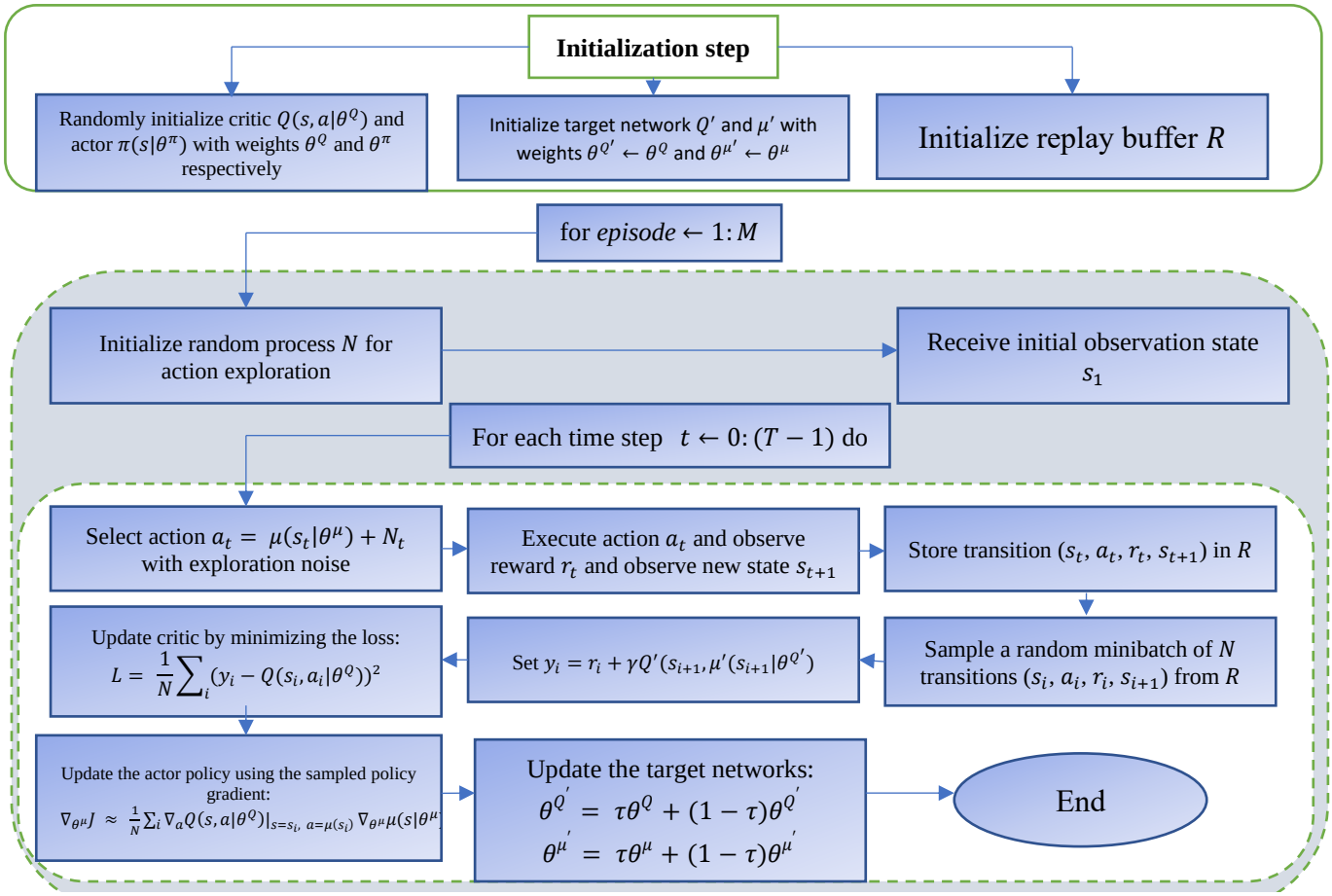

Figure 3. The DDPG Algorithm

summarized in flow chart figure 3.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 08, August-2021

E. DDPG parameter setting

The Algorithm needs to give action command for each generating unit that will satisfy all the constraints and meet the objective. The model is initialized by randomly selected power distribution coefficients for generator one to six as in table 2 below shows the initial operation parameters and table 3 shows the DDPG Algorithm parameter setting.

Table 2. Generating units' initial operation values

Symbol	Parameter	Value
g_1	Generating unit 1 power distribution coefficient	0.8
g_2	Generating unit 2 power distribution coefficient	0.6
g_3	Generating unit 3 power distribution coefficient	0.7
g_4	Generating unit 4 power distribution coefficient	0.8
g_5	Generating unit 5 power distribution coefficient	0.4
g_6	Generating unit 6 power distribution coefficient	0.6

Table 3. DDPG algorithm parameter setting

Parameter	Value
Target smooth factor	0.001
Experience buffer length	1000000
Discount factor	0.99
Minibatch size	32
Actor learning rate	0.0001
Critic learning rate	0.001

IV. EXAMPLE ANALYSIS

A. Simulation Environment

The GENCO mathematical model was developed in Simulink environment, constituted of an RL Agent block, Reward calculation subsystem and observation subsystem. The Deep reinforcement learning based on Deep Deterministic algorithm was created with the architecture explained in figure 1 and figure 2 for actor and critic respectively. The implementation was done by the help of deep designer app of MATLAB r2020b version. The setting parameters for the neural network architecture are tabulated in table 1.

Data used for training and testing the model

During training, the random time series data generator was formulated, this is to ensure good generalization of final result also it serves the purpose of large dataset. The standard IEEE 118-bus system data from [32] were utilized to test the trained RL agent. A single GENCO having six generating units of the 54 thermal units in the IEEE 118-bus test system and the generating units' data are given in table 4.

The input data to the model were 24-hour (day ahead) the time series predicted market clearing price and the predicted spot market allocation for the GENCO data as plotted in figure 4 and figure 5 respectively. The GENCO had six generating units each with different operation characteristics shown in the table 4.

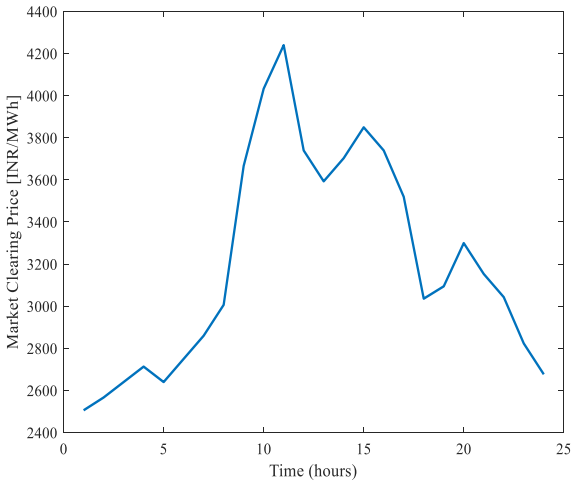

Figure 5. Predicted Market Clearing Price

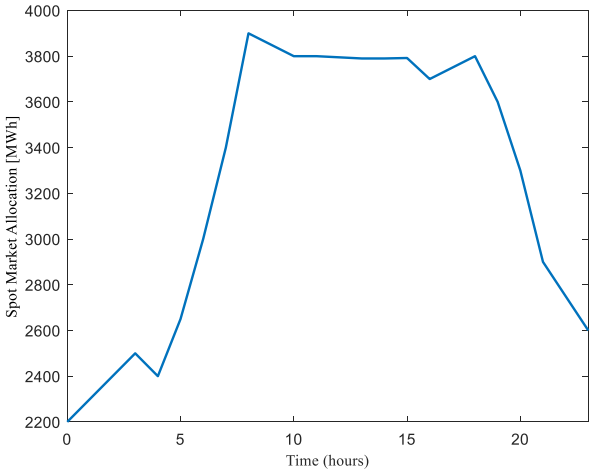

Figure 4. Predicted spot market allocation for the GENCO.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 08, August-2021

Table 4. Generating units' data

Unit Code	P_i^{min} [MW]	P_i^{max} [MW]	Capacity [MW]	a [INR/h] (x73.12)	b [INR/MWh] (x73.12)	C [INR/MWh^2] (x73.12)	MUT [hrs]	MDT [hrs]	RU [MW]	RD [MW]	HSC [INR/h] (x73.12)	CSC [INR/h] (x73.12)	CShr [hrs]
g1	100	420	840	128.32	16.68	0.0212	10	10	210	210	250	500	20
g2	100	300	2400	13.56	25.78	0.0218	8	8	150	150	110	110	16
g3	50	250	500	56.00	24.66	0.0048	8	8	125	125	100	200	16
g4	50	200	200	13.56	25.78	0.0218	8	8	100	100	400	800	16
g5	25	100	300	20.30	35.64	0.0256	5	5	50	50	50	100	10
g6	25	50	100	117.62	45.88	0.0195	2	2	25	25	45	90	4
Total			4340										

B. Results and Discussion

RL Agent training results

According to the algorithm, the model was trained for 150 episodes and each episode had 2400 steps. Each step returned a reward value which was summed to obtain an overall episode reward. Figure 5 shows a plot of episode reward against the episode number. The training was targeted to achieve at least average reward of 6300 for better results.

Testing the trained model

Upon successiful training of the Agent, the trained agent is applied in offline simulation to verify the agent's capability. In this scenario, the market clearing price and allocated energy are acting as inputs to the agent, and the agent gives optimal schedule

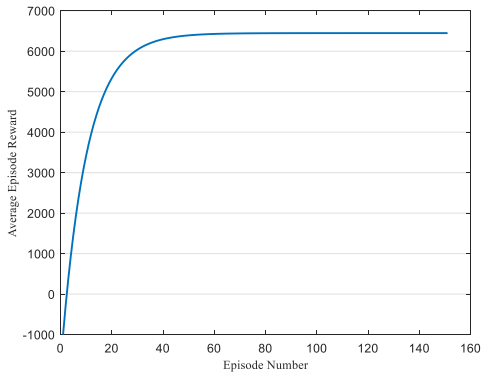

Figure 6. DDPG Average episode reward

To justify the results, trained agent was run in tree cases;

Case 1; the model was trained to meet the expected/ predicted spot market allocation.

case 2; all the generating units were fixed to generate their maximum capacity while generating unit two was optimized to minimize the operation cost.

case 3; the model was trained to find optimal bidding without fixing any of the generating units.

Table 5, is showing the amount of power assigned for each generating unit for each hour in for case 1, case 2 and case 3.

Table 6, summarizes the results obtained under three cases; case 1 had the highest operating cost as compared to case 2 and 3, this is because large amount of energy was being generated hence some units were operating under loss. The most optimal solution was under case 3, in which the operating cost was reduced and more power was being generated at instant that the clearing price is higher thus making more revenues. The case 3 made profit 1.5 times that in case 2 and 1.09 times that in case 2. The profit generated is increasing with increased energy price provided that the generator operation cost didn't reach its optimal value of operation. This was shown by the unit g_2, as compared to other units, its generation was following the MCP nature while others were constant. i.e higher generation was achieved at higher market clearing price provided that the total generated power doesn't exceed the spot market allocation. This can be proved by considering figure 4, figure 5 and the table 5. At the time between 7 to 19 hours, the market clearing price was high, this made the generated power to increase (as shown in table 5) in similar fashion as that of figure 4.

Table 5. Total generated profit

Case No.	Operation Cost (x10^8) [INR]	Total Revenue (x10^8) [INR]	Profit (x10^8) [INR]	Energy not supplied [MWh]
1	2.792	3.580	0.788	0.00
2	1.588	2.677	1.089	19010.00
3	1.783	2.968	1.185	12820.00

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 08, August-2021

Table 6. Amount of generated power by each generating unit at each time period of 24 hours

Time [Hrs]	g_1 (x840) [MW]			g_2 (x2400) [MW]			g_3 (x500) [MW]			g_4 (x200) [MW]			g_5 (x300) [MW]			g_6 (x100) [MW]		
	Case1	Case2	Case3	Case1	Case2	Case3	Case1	Case2	Case3	Case1	Case2	Case3	Case1	Case2	Case3	Case1	Case2	Case3
1	1	1	0.91	0.27	0.22	0.22	1	1	1	1	1	1	1	1	0.98	1	1	1
2	1	1	0.93	0.31	0.23	0.24	1	1	1	1	1	1	1	1	0.97	1	1	1
3	1	1	0.96	0.35	0.24	0.27	1	1	1	1	1	1	1	1	0.98	1	1	1
4	1	1	0.93	0.31	0.23	0.24	1	1	1	1	1	1	1	1	0.99	1	1	1
5	1	1	0.97	0.42	0.25	0.28	1	1	1	1	1	1	1	1	0.99	1	1	1
6	1	1	0.98	0.56	0.27	0.31	1	1	1	1	1	1	1	1	0.99	1	1	1
7	1	1	0.99	0.73	0.30	0.34	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	0.94	0.42	0.44	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	0.92	0.48	0.50	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	0.89	0.52	0.54	1	1	1	1	1	1	1	1	1	1	1	1
11	1	1	1	0.89	0.43	0.45	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	1	0.89	0.40	0.43	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	0.89	0.42	0.45	1	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	0.89	0.45	0.47	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	0.89	0.43	0.45	1	1	1	1	1	1	1	1	1	1	1	1
16	1	1	1	0.85	0.39	0.41	1	1	1	1	1	1	1	1	1	1	1	1
17	1	1	0.99	0.87	0.30	0.34	1	1	1	1	1	1	1	1	0.99	1	1	1
18	1	1	0.99	0.89	0.31	0.35	1	1	1	1	1	1	1	1	0.99	1	1	1
19	1	1	0.99	0.81	0.35	0.38	1	1	1	1	1	1	1	1	0.98	1	1	1
20	1	1	0.99	0.69	0.32	0.36	1	1	1	1	1	1	1	1	0.99	1	1	1
21	1	1	0.99	0.52	0.30	0.34	1	1	1	1	1	1	1	1	0.99	1	1	1
22	1	1	0.98	0.46	0.26	0.30	1	1	1	1	1	1	1	1	0.98	1	1	1
23	1	1	0.95	0.39	0.24	0.25	1	1	1	1	1	1	1	1	0.98	1	1	1
24	1	1	0.89	0.33	0.21	0.21	1	1	1	1	1	1	1	1	0.98	1	1	1

V. CONCLUSION

In this paper, deep reinforcement learning was used to find optimal scheduling to solve the profit-based generation unit scheduling of the GENCO operating in deregulated electricity market. The Deep Deterministic Policy Gradient algorithm is used to train the agent. The GENCO is assumed to operate under pool market without bilateral contacts of power supply between the GENCO and consumers. The important data input are; predicted spot market allocation for the GENCO and the market clearing price 24 hours (day ahead) of time. The method can also be applied in very complicated scenarios where there is larger number of constraints and many generation units.

REFERENCES

[1] K. V. N. P. Kumar *et al.*, "Expanding the Ambit of Ancillary Services in India - Implementation and Challenges," *2018 20th Natl. Power Syst. Conf. NPSC 2018*, 2018, doi: 10.1109/NPSC.2018.8771784.

[2] D. K. Mishra, T. K. Panigrahi, A. Mohanty, P. K. Ray, and M. Viswavandya, "Design and Analysis of Renewable Energy based Generation Control in a Restructured Power System," *Proc. 2018 IEEE Int. Conf. Power Electron. Drives Energy Syst. PEDES 2018*, 2018, doi: 10.1109/PEDES.2018.8707753.

[3] Y. R. Prajapati, V. N. Kamat, J. J. Patel, and B. Parekh, "Impact of grid connected solar power on load frequency control in restructured power system," *2017 Innov. Power Adv. Comput. Technol. i-PACT 2017*, vol. 2017-Janua, pp. 1–5, 2017, doi: 10.1109/IPACT.2017.8245122.

[4] J. Li, Z. Li, and Y. Wang, "Optimal bidding strategy for day-ahead power market," *2015 North Am. Power Symp. NAPS 2015*, pp. 1–6, 2015, doi: 10.1109/NAPS.2015.7335133.

[5] K. Choudhary, R. Kumar, D. Upadhyay, and B. Singh, "Optimal Power Flow Based Economic Generation Scheduling in Day-ahead Power Market," *Int. J. Appl. Power Eng.*, vol. 6, no. 3, p. 124, 2017, doi: 10.11591/ijape.v6.i3.pp124-134.

[6] K. Lakshmi and S. Vasantharathna, "A profit based unit commitment problem in deregulated power markets," *2009 Int. Conf. Power Syst. ICPS '09*, no. 2, pp. 25–30, 2009, doi: 10.1109/ICPWS.2009.5442772.

[7] P. Che and L. Tang, "Stochastic unit commitment with CO 2 emission trading in the deregulated market environment," *Asia-Pacific Power Energy Eng. Conf. APPEEC*, no. 70728001, pp. 10–13, 2010, doi: 10.1109/APPEEC.2010.5448754.

[8] "Unit Commitment in a Power Generation System Using a modified Improved-Dynamic Programming Henry U gochukwu Ukwu Reza Sirjani," pp. 103–107, 2016.

[9] A. Bhardwaj, V. K. Kamboj, V. K. Shukla, B. Singh, and P. Khurana, "Unit commitment in electrical power system - A literature review," *2012 IEEE Int. Power Eng. Optim. Conf. PEOCO 2012 - Conf. Proc.*, no. June, pp. 275–280, 2012, doi: 10.1109/PEOCO.2012.6230874.

[10] S. V. Tade, V. N. Ghate, S. Q. Mulla, and M. N. Kalgunde, "Application of Dynamic Programming Algorithm for Thermal Unit Commitment with Wind Power," *Proc. - 2018 IEEE Glob. Conf. Wirel. Comput. Networking, GCWCN 2018*, no. 2, pp. 182–186, 2019, doi: 10.1109/GCWCN.2018.8668612.

[11] C. Su, C. Cheng, and P. Wang, "An MILP Model for Short-Term Peak Shaving Operation of Cascaded Hydropower Plants Considering Unit Commitment," *Proc. - 2018 IEEE Int. Conf. Environ. Electr. Eng. 2018 IEEE Ind. Commer. Power Syst. Eur. EEEIC/I CPS Eur. 2018*, no. 978, pp. 1–6, 2018, doi: 10.1109/EEEIC.2018.8494460.

[12] V. Arora and S. Chanana, "Solution to unit commitment problem using Lagrangian relaxation and Mendel's GA method," *Int. Conf. Emerg. Trends Electr. Electron. Sustain. Energy Syst. ICETEESES 2016*, pp. 126–129, 2016, doi: 10.1109/ICETEESES.2016.7581372.

[13] S. Siva Sakthi, R. K. Santhi, N. Murali Krishnan, S. Ganesan, and S. Subramanian, "Wind Integrated Thermal Unit Commitment Solution using Grey Wolf Optimizer," *Int. J. Electr. Comput. Eng.*, vol. 7, no. 5, pp. 2309–2320, 2017, doi: 10.11591/ijece.v7i5.pp2309-2320.

[14] Y. Zhang, "A Novel hybrid immune particle swarm optimization algorithm for unit commitment considering the environmental cost," *Proc. - 2019 Int. Conf. Smart Grid Electr. Autom. ICSGEA 2019*, pp. 54–58, 2019, doi: 10.1109/ICSGEA.2019.00021.

[15] M. Jabri, H. Aloui, and H. A. Almuzaini, "Fuzzy logic lagrangian relaxation selection method for the solution of unit commitment problem," *2019 8th Int. Conf. Model. Simul. Appl. Optim. ICMSAO 2019*, pp. 2019–2022, 2019, doi: 10.1109/ICMSAO.2019.8880285.

[16] B. Hu, Y. Gong, and C. Y. Chung, "Flexible Robust Unit Commitment Considering Subhourly Wind Power Ramp Behaviors," *2019 IEEE Can. Conf. Electr. Comput. Eng. CCECE 2019*, pp. 30–33, 2019, doi: 10.1109/CCECE.2019.8861590.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 08, August-2021

[17] Anamika and N. Kumar, "Market Clearing Price prediction using ANN in Indian Electricity Markets," *2016 Int. Conf. Energy Effic. Technol. Sustain. ICEETS 2016*, pp. 454–458, 2016, doi: 10.1109/ICEETS.2016.7583797.

[18] L. Ding and Q. Ge, "Electricity Market Clearing Price Forecast Based on Adaptive Kalman Filter," *ICCAIS 2018 - 7th Int. Conf. Control. Autom. Inf. Sci.*, no. Iccais, pp. 417–421, 2018, doi: 10.1109/ICCAIS.2018.8570534.

[19] H. Iwasaki and A. Okuyama, "Development of a reference signal self-organizing control system based on deep reinforcement learning," *2021 IEEE Int. Conf. Mechatronics, ICM 2021*, pp. 1–5, 2021, doi: 10.1109/ICM46511.2021.9385676.

[20] E. Xing and B. Cai, "Delivery Route Optimization Based on Deep Reinforcement Learning," *Proc. - 2020 2nd Int. Conf. Mach. Learn. Big Data Bus. Intell. MLBDBI 2020*, pp. 334–338, 2020, doi: 10.1109/MLBDBI51377.2020.00071.

[21] Y. Long and H. He, "Robot path planning based on deep reinforcement learning," *2020 IEEE Conf. Telecommun. Opt. Comput. Sci. TOCS 2020*, pp. 151–154, 2020, doi: 10.1109/TOCS50858.2020.9339752.

[22] H. Abdi, "Profit-based unit commitment problem: A review of models, methods, challenges, and future directions," *Renew. Sustain. Energy Rev.*, vol. 138, no. October, p. 110504, 2021, doi: 10.1016/j.rser.2020.110504.

[23] A. K. Bikeri, C. M. Muriithi, and P. K. Kihato, "A review of unit commitment in deregulated electricity markets," *Proc. Sustain. Res. Innov. Conf.*, no. May, pp. 9–13, 2015, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.840.1938&rep=rep1&type=pdf.

[24] E. F. Morales and J. H. Zaragoza, "An introduction to reinforcement learning," *Decis. Theory Model. Appl. Artif. Intell. Concepts Solut.*, pp. 63–80, 2011, doi: 10.4018/978-1-60960-165-2.ch004.

[25] D. Arora, M. Garg, and M. Gupta, "Diving deep in Deep Convolutional Neural Network," *Proc. - IEEE 2020 2nd Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2020*, pp. 749–751, 2020, doi: 10.1109/ICACCCN51052.2020.9362907.

[26] S. Kamiş and D. Goularas, "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data," *Proc. - 2019 Int. Conf. Deep Learn. Mach. Learn. Emerg. Appl. Deep. 2019*, pp. 12–17, 2019, doi: 10.1109/Deep-ML.2019.00011.

[27] J. C. Jesus, J. A. Bottega, M. A. S. L. Cuadros, and D. F. T. Gamarra, "Deep Deterministic Policy Gradient for Navigation of Mobile Robots in Simulated Environments," 2019.

[28] M. Zhang, Y. Zhang, Z. Gao, and X. He, "An Improved DDPG and Its Application Based on the Double-Layer BP Neural Network," vol. 8, 2020, doi: 10.1109/ACCESS.2020.3020590.

[29] X. Guo *et al.*, "A Novel User Selection Massive MIMO Scheduling Algorithm via Real Time DDPG," 2020.

[30] C. Chu, K. Takahashi, and M. Hashimoto, "Comparison of Deep Reinforcement Learning Algorithms in a Robot Manipulator Control Application," pp. 284–287, 2020, doi: 10.1109/IS3C50286.2020.00080.

[31] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," *31st Int. Conf. Mach. Learn. ICML 2014*, vol. 1, pp. 605–619, 2014.

[32] B. Pmin, P. Qmin, Q. Min, and P. Coef, "Table I Market price Table II Fossil Unit Data Table III Combined Cycle Unit Data Table IV Combined Cycle Unit Configuration Data," pp. 1–4.