energies

MDPI

Article

Novel Energy Trading System Based on Deep-Reinforcement Learning in Microgrids

Seongwoo Lee, Joonho Seon, Chanuk Kyeong, Soohyun Kim, Youngghyu Sun and Jinyoung Kim *

Department of Electronic Convergence Engineering, University of Kwangwoon, Seoul 01897, Korea; swoo1467@kw.ac.kr (S.L.); dimlight13@kw.ac.kr (J.S.); rudcksdnr@kw.ac.kr (C.K.); kimsoogus@kw.ac.kr (S.K.); yakrkr@kw.ac.kr (Y.S.)

* Correspondence: jinyoung@kw.ac.kr; Tel.: +82-02-940-5567

Abstract: Inefficiencies in energy trading systems of microgrids are mainly caused by uncertainty in non-stationary operating environments. The problem of uncertainty can be mitigated by analyzing patterns of primary operation parameters and their corresponding actions. In this paper, a novel energy trading system based on a double deep Q-networks (DDQN) algorithm and a double Kelly strategy is proposed for improving profits while reducing dependence on the main grid in the microgrid systems. The DDQN algorithm is proposed in order to select optimized action for improving energy transactions. Additionally, the double Kelly strategy is employed to control the microgrid's energy trading quantity for producing long-term profits. From the simulation results, it is confirmed that the proposed strategies can achieve a significant improvement in the total profits and independence from the main grid via optimized energy transactions.

Keywords: microgrid; energy transaction; energy self-sufficient systems; double deep Q-networks (DDQN); double Kelly strategy

check for updates

Citation: Lee, S.; Seon, J.; Kyeong, C.; Kim, S.; Sun, Y.; Kim, J. Novel Energy Trading System Based on Deep-Reinforcement Learning in Microgrids. *Energies* **2021**, *14*, 5515. https://doi.org/10.3390/en14175515

Academic Editor: Mohamed Benbouzid

Received: 16 August 2021
Accepted: 1 September 2021
Published: 3 September 2021

1. Introduction

Recently, microgrid (MG) systems have been widely established for distributed power grids [1,2]. Compared with conventional grid systems, energy transmission losses and carbon emissions can be minimized by the MG [3]. The MG has been typically designed for the self-sufficiency of energy by operating energy management and transactions independently from the main grid [4–6]. Therefore, the MG can relieve disasters such as cyber-physical attacks and power outages [4]. In an energy trading system (ETS), the MG acts as a prosumer and interacts to maximize social welfare through a distributed power system of an electricity trade model [6,7]. To operate off-grid, the ETS requires the forecasting of demand and supply requirements and the balancing of the possession of energy [5]. In addition, profits should be guaranteed to prove the economic feasibility of the ETS.

Several studies have been conducted for a variety of operating environments on the energy independence and profitability of the MG system [8–11]. In [8], multi-MG interconnection and policies have been proposed for independent energy management, which eventually saves social energy trading costs and improves reliability. An optimized strategy for each MG has been recommended for an efficient grid-to-grid (G2G)-based energy trading method [9]. In [10], the interconnected multi-MG energy trading model has been proposed for the optimal energy scheduling of MGs by improving the energy efficiency of the system. Independence of the energy from the main grid can be achieved via interdependence between developed and developing MGs when the G2G-based energy trade is formed [11]. In [12], temporal complementarity between supply and demand in energy trade has been proposed for the entire community of energy self-sufficiency. For an efficient ETS, fluent switching between consumer and seller can be a crucial factor for variant operating situations. Another important aspect of the ETS is improving profits in a balanced way. In order to improve the profits via energy trading, it is necessary to predict and utilize information about each MG. In order to predict energy surplus or deficit, the ETS can be modeled as a resilience system [13]. For maximizing the benefits to consumers, optimal transacted quantities have been determined by considering uncertainties and coordinating day-ahead optimal scheduling [14]. In addition, several algorithms have been proposed to manage the MG for profits and balanced social welfare [15]. A compute clearing price and volume (CCPV) algorithm has been introduced to maximize the equilibrium quantity of energy trading for maximizing profits [15]. Additionally, an envy-free division (EFD) algorithm has been proposed to increase social welfare by allowing agents to equally share opportunities for profits [15]. In [16], a multi-bilateral economic dispatch formulation has been proposed based on optimal power flow to maximize community welfare while avoiding critical grid conditions.

In order to optimize energy trading, several schemes have been proposed and analyzed [17–20]. Reinforcement learning (RL) based algorithms have been proposed for the improvement in decision-making for the energy management of MG [17–20]. For operating any bounded utility of stochastic character, a linear-inaction scheme has been proposed to maximize average revenue through the incorporation of the RL algorithm and game theory [18]. The RL algorithm has been used to reduce the supply-demand mismatch problem for optimizing the energy-independent model [19]. Considering practical energy trading environments, the MG energy management has been modeled by the deep-RL algorithm for solving the problem of random variations in both supply and demand [20]. However, the presented systems have generally been focused on short-term efficiency for constructing a real-time trading system without considering the long-term perspective. Furthermore, due to simulation in stationary environments, they have applicability limitations to the realistic ETS.

In this paper, a novel G2G-based ETS is proposed for a self-sufficient energy system which can manage independence on the main grid as well as profits by lower trading costs [11,21]. Utilizing a double Kelly strategy, the optimized quantity of energy trading can be computed in the uncertainty of the future for each MG. The ETS model is optimized by a double deep Q-networks (DDQN) to improve the performance of the ETS. Therefore, compared with the conventional approaches, contributions of the proposed system can be summarized in the following three aspects:

- Considering the quantity of energy trading, conventional studies have been largely determined by static trading systems. Unlike conventional studies, we employed the double Kelly strategy, considering dynamic determinations and flexible energy transactions. The proposed ETS can improve the utility of energy resources and the activation of energy transactions;
- Conventional studies have typically been simulated in stationary environments. However, the proposed system is simulated in non-stationary environments in order to reflect a more realistic market scenario, considering the changeable patterns of the quantity of energy trading, surplus energy, energy prices, energy production by weather, and monthly demand;
- Considering the ETS, it is shown that most of the existing schemes have typically focused on profitability rather than independence. The proposed system has paid attention to both, simultaneously achieving profitability and independence by the DDQN algorithm and the double Kelly strategy.

The rest of this paper is organized as follows. In Section 2, the novel energy trading system model of the MG is introduced. In Section 3, simulation results of the energy transaction are presented, and the performance of the proposed system is analyzed. Finally, conclusions are drawn in Section 4.

2. Novel Energy Trading System Model

The configuration of the G2G-based ETS in this paper is shown in Figure 1. In the MG, solar energy resources are used as the power source. Insufficient energy quantity can be supplied by the main grid. The energy transactions are performed by determining the optimal quantity of energy trading based on the double Kelly strategy. The MGs are sought to reach a consensus by acting to maximize their respective interests and thus benefitting entire system.

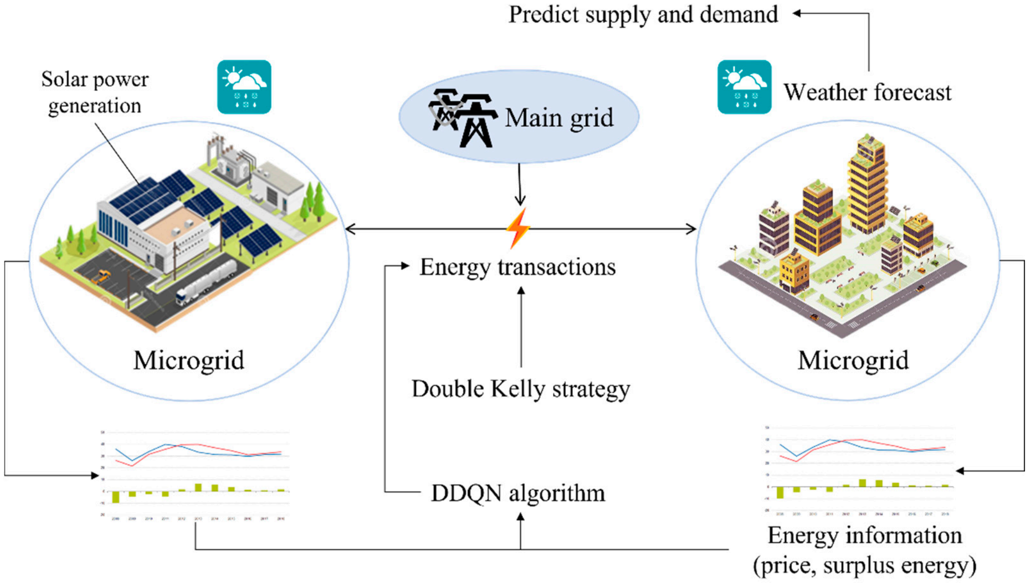


Figure 1. The proposed energy trading scheme based on the G2G model.

2.1. Double Kelly Strategy

The Kelly strategy has been known to pursue long-term profits and prepare for the uncertainty of the future [22]. The double Kelly strategy is a refined version of the Kelly strategy for the activation of the energy trading market. The limitation of analyzing the risky transactions via the current state can be solved by using the double Kelly strategy. As a result, optimal transactions are conducted to minimize risk by considering the loss of the transaction. In addition, the double Kelly strategy sees the transaction proceed flexibly according to the situation, on a case-by-case basis. Applying the double Kelly strategy to the energy trading model, the energy is regarded as capital from the Kelly function [22]. The winning probability is given by

$$0 \leq p_w \leq 1, \tag{1}$$

where p_w denotes the winning probability. In the proposed ETS, p_w is set by the energy production rate related to weather [23]. The maximum value of p_w can be achieved when energy production is optimized. A profit and loss can be represented by

$$r_W = \mathrm{B}_{exp}^{(t)} + 1, \tag{2}$$

$$\mathrm{B}_{exp}^{(t)} = S^{(t)} p_{change}, \tag{3}$$

where r_W is compensation for good transactions, $\mathrm{B}_{exp}^{(t)}$ is the expected benefits in energy transactions at time step t, $S^{(t)}$ is stored energy and p_{change} is expected price change. The value of B_{exp} can be found by predicting the price and profits through energy trading using the transition probability of the weather.

Assuming a certain probability of profit and loss with initial energy quantity set to be E_0, the energy quantity of E_1 after the transaction is given by

$$\mathrm{E}_1 = \mathrm{E}_0(1 + \mathrm{K}r_W)^{W_1}(1 + \mathrm{K}r_L)^{1-W_1}, \tag{4}$$

$$\mathrm{E}_{1,win} = \mathrm{E}_0(1+\mathrm{K}r_W), \tag{5}$$

$$\mathrm{E}_{1,lose} = \mathrm{E}_0(1+\mathrm{K}r_L), \tag{6}$$

where K is a trading ratio, W is a random variable which is 1 with p_w and 0 with $1-p_w$, and r_L is a loss in energy transactions. The case of E_2 after two transactions can be expressed by

$$\mathrm{E}_2 = \mathrm{E}_1(1+\mathrm{K}r_W)^{W_2}(1+\mathrm{K}r_L)^{1-W_2} = \mathrm{E}_0(1+\mathrm{K}r_W)^{W_1+W_2}(1+\mathrm{K}r_L)^{2-W_1-W_2}. \tag{7}$$

The energy E_n after n transactions can be written as

$$\mathrm{E}_n = \mathrm{E}_0(1+\mathrm{K}r_W)^{(\Sigma_{i=1}^{n} W_i)}(1+\mathrm{K}r_L)^{(n-\Sigma_{i=1}^{n} W_i)}. \tag{8}$$

When $W_{N,win}$ is the number of high performances with N times trade, it can be rewritten as

$$\mathrm{E}_n = \mathrm{E}_0(1+\mathrm{K}r_W)^{W_{N,win}}(1+\mathrm{K}r_L)^{n-W_{N,win}}. \tag{9}$$

where $W_{N,win}$ is the number of good performances out of a total of n trades. Since $W_{N,win}$ is a variable following a binomial distribution, the average np_w is obtained by applying the law of large numbers under the assumption that N is large enough. Then, the energy E_n after n transactions can be modified by

$$\mathrm{E}_n = \mathrm{E}_0(1+\mathrm{K}r_W)^{np_w}(1+\mathrm{K}r_L)^{n(1-p_w)}, \tag{10}$$

The K_{opt} at the point where the $\mathrm{E}_n/\mathrm{E}_0$ value is maximized can be given by

$$\begin{aligned}\mathrm{K}_{opt} &= \underset{\mathrm{K}}{\operatorname{argmax}}\left(\frac{\mathrm{E}_n}{\mathrm{E}_0}\right) = \underset{\mathrm{K}}{\operatorname{argmax}}\left(ln\left(\frac{\mathrm{E}_n}{\mathrm{E}_0}\right)\right) \\ &= \underset{\mathrm{K}}{\operatorname{argmax}}\left(ln\left((1+\mathrm{K}r_W)^{np_w}(1+\mathrm{K}r_L)^{n(1-p_w)}\right)\right)\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{d}{dl}\left(ln((1+\mathrm{K}r_W)^{np_w}(1+\mathrm{K}r_L)^{n(1-p_w)})\right) &= \frac{np_w r_W}{1+\mathrm{K}r_W} + \frac{n(1-p_w)r_L}{1+\mathrm{K}r_L} \\ &= \frac{n(p_w r_W + r_L + \mathrm{K}r_W r_L - p_w r_L)}{(1+\mathrm{K}r_W)(1+\mathrm{K}r_L)} = 0,\end{aligned} \tag{12}$$

$$\mathrm{K}_{opt} = -\frac{p_w}{r_L} - \frac{1-p_w}{r_W}, \tag{13}$$

In general, since $r_L = -1$, it can be expressed as

$$\mathrm{K}_{opt} = p_w - \frac{1-p_w}{r_W}, \tag{14}$$

The tradeable energy for transactions which can be derived from the double Kelly strategy can be described by

$$\mathrm{K}_{double} = 2\mathrm{K}_{opt}, \tag{15}$$

$$T_i^{(t)} = \mathrm{K}_{double} b_i^{(t)}, \tag{16}$$

where K_{double} denotes tradeable energy quantity factor using double Kelly strategy, $T_i^{(t)}$ is the tradeable quantity of energy of MG_i at time step t and $b_i^{(t)}$ is the surplus energy of MG_i at time step t. In Equation (16), the long-term profits can be obtained by determining an appropriate quantity of energy trading for each time step.

2.2. Reinforcement Learning for Optimized Energy Trading System

2.2.1. Q-Learning Algorithm

A Q-learning algorithm is employed to build an energy trading system. It is a well-known model-free RL algorithm for solving discrete spatial problems [24]. In the Q-learning algorithm, the optimal policy is to maximize the total reward by successive states. When a specific action a is performed in a certain state s for the Q-learning algorithm, the Q-value is used to determine the value of the action. The Q-value is then determined by the following equations

$$Q_{t+1}(s_t, a_t) = (1 - lr)Q_t(s_t, a_t) + lr(r_{t+1} + \gamma maxQ(s_{t+1}, a_{t+1})), \tag{17}$$

$$X_t^Q = r_{t+1} + \gamma maxQ(s_{t+1}, a_{t+1}), \tag{18}$$

where lr is the learning rate of Q-function, s_t is the state at time t, a_t is the action at time step t, r_{t+1} is the reward at time step $t + 1$ and γ is the discount factor. The Q-value converges to the optimal Q-value to derive the optimal policy by repeating Equation (17). The sampling of the target network in Equation (18) is required for convergence.

The Q-learning algorithm has been used to converge optimal policy to improve system performance [25,26]. However, there are a couple of challenging problems with the optimal policy. One is the memory shortage problem [27], and the other is overestimation [28]. Deep Q-networks (DQN) algorithm has been proposed to solve the memory shortage problem [27]. However, the overestimation problem still exists in the DQN algorithm. Therefore, deep-RL with a double Q-learning algorithm named DDQN algorithm has been proposed to solve these two problems simultaneously [28].

2.2.2. DDQN Algorithm

The DDQN algorithm is approximated by adding a neural network to overcome the memory shortage problem; the overestimation problem can be solved by adding an additional neural network. The Q-value in the DDQN algorithm is determined by the following equations

$$Q_{\mathrm{A}+1}(s_t, a_t) = (1 - lr)Q_A(s_t, a_t) + lr(r_{t+1} + \gamma Q_B(s_{t+1}, argmaxQ_A(s_{t+1}, a_{t+1})), \tag{19}$$

$$\mathrm{X}_t^{DDQN} = r_{t+1} + \gamma Q_B(s_{t+1}, argmaxQ_A(s_{t+1}, a_{t+1})), \tag{20}$$

where Q_A and Q_B are primary and target networks, respectively. In Equation (19), the expression $maxQ(s_{t+1},\ a_{t+1})$ from Equation (17) is divided into selection and evaluation parts [29]. In the decision-making process, the DDQN algorithm is used to select an action and evaluate the selected action. In the DQN algorithm, the selection and evaluation parts are used by only one network. However, the overestimation problem can be caused by the DQN algorithm with only one network. Dividing one network into two networks is one of the methods which can be employed to overcome the overestimation problem. As shown in Equation (20), networks are separated in the DQN algorithm to reach an optimal value. The Q_A and Q_B are used to select actions and to evaluate the Q-value from selected actions, respectively. As the action space increases, the estimation error of the Q-value can be increased in the DQN algorithm. On the other hand, the estimation error of the DDQN algorithm can be underestimated rather than overestimated due to low variance in comparison to the DQN algorithm [28]. Nevertheless, the DDQN algorithm is used for the advantages of error estimation and stable learning in the proposed ETS.

2.3. Proposed Energy System

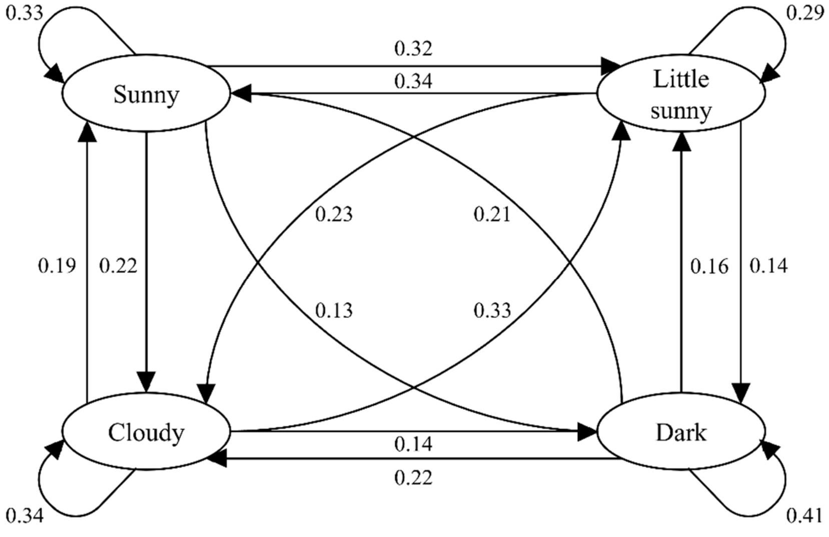
2.3.1. Data Settings

The monthly consumption patterns and weather transition probabilities have been analyzed to generate reliable one-year data of MG. The average rate of change in consumption, which varies monthly, is shown in Table 1 [30].

Table 1. The rate of the monthly change in average energy demand.

Month	**January**	**February**	**March**	**April**	**May**	**June**
Monthly demand change ratio	+19%	+19%	−2%	−4%	−14%	−11%
Month	**July**	**August**	**September**	**October**	**November**	**December**
Monthly demand change ratio	−10%	+7%	−2%	−10%	+2%	+6%

In Figure 2, transition probability by weather is calculated by analyzing 10 years of weather data from the Korea Meteorological Administration. The energy generation efficiency, according to weather state, is averaged and used to create data.



$\eta_{sunny} = 1,\ \eta_{little\ sunny} = 0.8,\ \eta_{cloudy} = 0.7,\ \eta_{dark} = 0.6$

Figure 2. Transition probability and energy production efficiency of weather.

In order to make a more realistic proposed model, the demand, supply, weather, and efficiency of production are considered in the process of data settings. For the reliability of the data, one year's worth of data is generated with added variance.

2.3.2. Energy Transaction System Function Settings

The energy price is determined by the cost price of MGs and the price ratio. The energy price is given by

$$P_i^{(t)} = cost_{price}(1 + p_i), \tag{21}$$

$$p_i = \left(M^{(t)} + 1\right) / \left(\frac{1 + b_i^{(t)}}{W_i^{(t)}}\right), \tag{22}$$

where $cost_{price}$ denotes levelized cost of electricity (LCOE), p_i is price ratio of MG_i, $M^{(t)}$ is monthly demand variance by the MG's consumption patterns and $W_i^{(t)}$ is expected energy production efficiency by the weather of MG_i at time step t. The profits of energy transactions can be expressed as

$$\eta = B_{sale} + B_{purchase}, \tag{23}$$

$$B_{sale} = B_{trading} + B_{storage}, \tag{24}$$

$$B_{purchase} = B_{margin}^{Maingrid} + B_{margin}^{trading}, \tag{25}$$

where η is energy transaction profit of MGs, B_{sale} is benefit of energy sales, $B_{purchase}$ is benefit of energy purchase, $B_{trading}$ is benefit of the energy trading from the seller's point of view, $B_{storage}$ is benefit of the energy storage for expected profit, $B_{margin}^{Maingrid}$ is benefit of margin compared to trading with the main grid and $B_{margin}^{trading}$ is benefit of the price margin of the transaction. The η is used to keep a balance of the profits between supplier and consumer. The high profits can be achieved in the MGs when they conduct a price-efficient transaction with appropriate energy. Then, the benefits from energy trading are given by

$$B_{trading} = p_i^{(t)} T_i^{(t)}, \tag{26}$$

$$B_{storage} = E_{i,remain}^{(t)} B_{exp}^{(t)}, \tag{27}$$

$$B_{margin}^{Maingrid} = \left(p_{maingrid} - p_{i,bid}^{(t)}\right) T_i^{(t)}, \tag{28}$$

$$B_{margin}^{trading} = \left(p_{trading}^{(t)} - p_{i,bid}^{(t)}\right) T_i^{(t)}, \tag{29}$$

where $E_{i,remain}^{(t)}$ is the remaining energy after MG's energy transaction at time step t, $p_{maingrid}$ is energy price of the main grid, $p_{i,bid}^{(t)}$ is the bid energy price of MG_i at time step t, and $p_{trading}^{(t)}$ is energy trading price of MG_i at time step t. The value of $p_{trading}^{(t)}$ is chosen in the middle of the values between supplier and consumer price of energy. The DDQN reward can be described as

$$r = \begin{cases} 0 & V_p^{(t)} < V_p^{thr}, V_d^{(t)} < V_d^{thr} \\ \alpha; & V_p^{(t)} < V_p^{thr}, V_d^{(t)} > V_d^{thr} \text{ or } V_p^{(t)} > V_p^{thr}, V_d^{(t)} < V_d^{thr} \\ 1; & V_p^{(t)} > V_p^{thr}, V_d^{(t)} > V_d^{thr} \end{cases} \tag{30}$$

where α is a constant value between 0 and 1, $V_p^{(t)}$ is profit at time step t, V_p^{thr} is a threshold of the profit, $V_d^{(t)}$ is dependency at time step t and V_d^{thr} is a threshold of dependency.

2.3.3. Energy Trading Scheme with the Double Kelly Strategy

The proposed energy trading scheme can be modeled by the double Kelly strategy and the DDQN for the improvement of the energy trading system. The flowchart of the proposed scheme is represented in Figure 3.

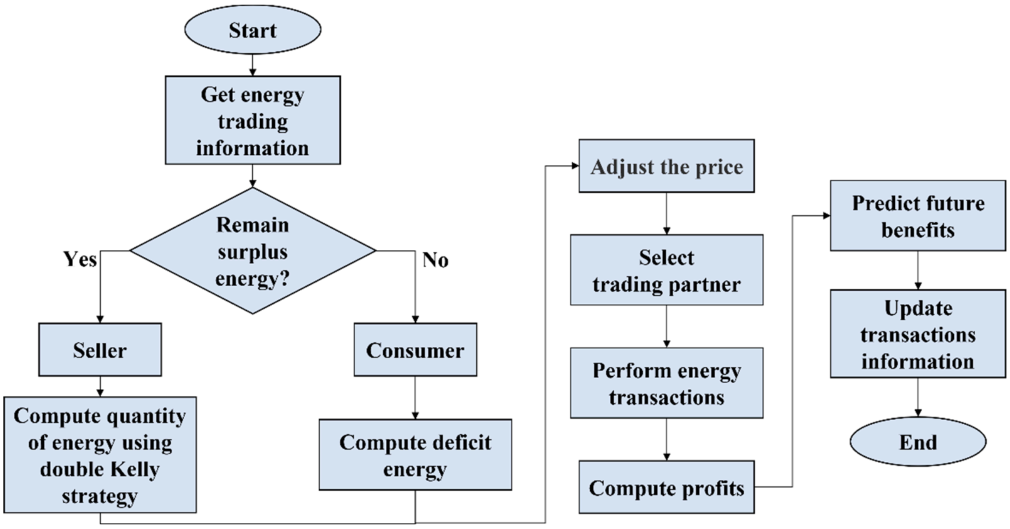


Figure 3. Flowchart of the proposed energy trading scheme.

The energy trading partners are chosen by the proposed energy trading scheme. The quantity of energy trading is determined by the double Kelly strategy. The proposed DDQN algorithm-based energy trading scheme is summarized in Algorithm 1.

Algorithm 1 Proposed DDQN algorithm-based energy trading scheme

```
Initialize agent parameters, DDQN networks ($Q_{evaluation}, Q_{target}$), buffer memory $B$
Initialize state parameters $state_i^{(t)} = \left[S_i^{(t)}, T_i^{(t)}, P_i^{(t)}, M^{(t)}, W_i^{(t)}\right]$
for $Episodes = 1, 2, \ldots, N$ do
    if $Episodes > Episodes_{Max}$ then
        Get initial $state_i^{(t)} = \left[S_i^{(t)}, T_i^{(t)}, P_i^{(t)}, M^{(t)}, W_i^{(t)}\right]$
        for $Steps\ t = 1, 2, \ldots, M$ do
            if $Steps < Steps_{Max}$ then
                Estimate $generation_i^{(t)}, demand_i^{(t)}$ from $W_i^{(t)}, M^{(t)}$
                Compute $P_i^{(t)}$ by (21) and (22)
                Compute $K_{i,double}^{(t)}$ by (1)–(15)
                Get $T_i^{(t)}$ by Equation (16)
                Get $state_i^{(t)} = \left[S_i^{(t)}, T_i^{(t)}, P_i^{(t)}, M^{(t)}, W_i^{(t)}\right]$
                Send $state_i^{(t)}$ to DDQN
                Remember memory in $B$
                if $B > B_{max}$ then
                    Replay memory and update target network
                end if
                Select action $a_i^{(t)}$ with $\varepsilon - greedy$ strategy and send to the environment
                Execute energy transactions by $a_i^{(t)}$
                Check $overflow$ and conduct transaction
                Compute $V_p^{(t)}\ and\ V_d^{(t)}$
                Compute reward $r^{(t)}$ by (23)–(30)
                Observe next state $state_i^{(t+1)}$ and reward $r^{(t)}$
                Store ($state_i^{(t)}, a_i^{(t)}, r^{(t)}, state_i^{(t+1)}$) and $done$
            end if
        end for
    end if
end for
```

3. Simulation Results

3.1. Settings in Agents, Data, and System Parameters

It is assumed that all agents have the capability of storing and producing energy. The agents are set up with four MGs and consist of two types of traders. The types of traders are composed of consumers and suppliers [31–33]. A consumer-type MG represents the MG which consumes energy most of the time while a supplier-type MG denotes the MG which produces a large trading energy quantity.

It is well known that energy demand and production are highly affected by seasonal data settings reflecting weather conditions [31–36]. In the simulations, their impacts have been taken into account for the justification of the proposed strategies by obtaining practical results in non-stationary operating environments.

In Table 2, hyperparameters used in the simulations are presented for the proposed ETS.

Table 2. List of hyperparameters and values.

Parameter	Value	Parameter	Value
Episodes	2500	Discount rate	0.99
Simulation step	365	Epsilon minimum	0.01
Learning rate	1×10^{-3}	Epsilon decay	0.999
Optimizer	Adam	Batch size	128
Layers	1D-CNN	Experience replay memory size	30,000

3.2. Performance Metrics

In this paper, the three kinds of performance metrics are employed to evaluate the proposed scheme. They include scores according to the DDQN algorithm, energy dependence on the main grid and total profits of MGs over a year. They are expected to have a critical impact on the improvement of profits and reducing main grid dependency in the MG with optimized ETS.

The final rewards, considering the main grid dependency and the profits through energy trading, are represented by the DDQN training scores. The proposed ETS is focused on balanced energy schedules which are considered alongside the profits and independence in the MGs. Therefore, an efficient ETS can be achieved when energy independence and profits from energy trading are balanced. A high score can be achieved by reducing the energy dependency and producing a highly lucrative level of the transaction. Therefore, the high score can bring about the result of activating the energy trading market in the MG.

As the paradigm shifts from centralized to distributed energy control, the ETS should act to the reduce the dependence on the main grid [36]. The off-grid MG has been spotlighted in order to ensure the stability of the entire society's energy supply system. In addition, independence on the main grid is required for quick restoration in the case of an accident in the power systems.

The total profits in energy transactions are a transparent measure in the ETS. In the energy market, increasing the quantity of energy trading and total profits of MGs have been recommended. The proposed ETS has been devised to pursue the profit of the entire MG rather than a monopoly for the long-term. Therefore, a high profit can be achieved by accurate prediction of the entire energy market and appropriate transactions in the ETS.

3.3. Simulation Results in Energy Trading Scenario

The improvements in independence on the main grid and profits are shown from the simulation results over a year-long period. The simulations were repeatedly operated until a plateau was reached due to the cost of the DDQN methods. In Figure 4, the three strategies (the proposed double Kelly strategy, EFD [15], and CCPV [15]) of the ETS are compared in terms of the DDQN learning score. A high score of convergences is revealed by the DDQN-based energy trading in the proposed strategy. An average score of 260 after episode 350 is shown by the proposed strategy. It is confirmed that a balance between the profits via energy trading and energy independence can be achieved by the proposed system. Similarly to the proposed strategy, an average score of 250 is shown in the CCPV-ETS. However, a low score is shown in the EFD-ETS because it implements energy trading free from competition.

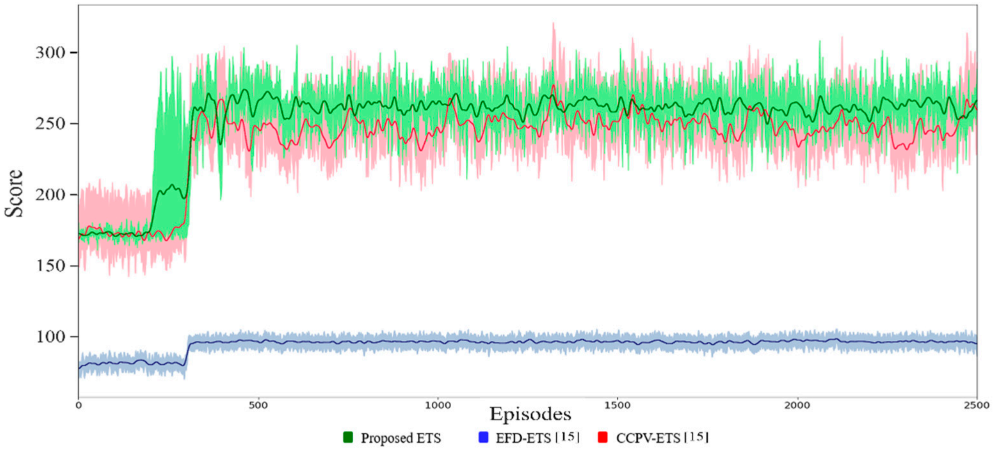
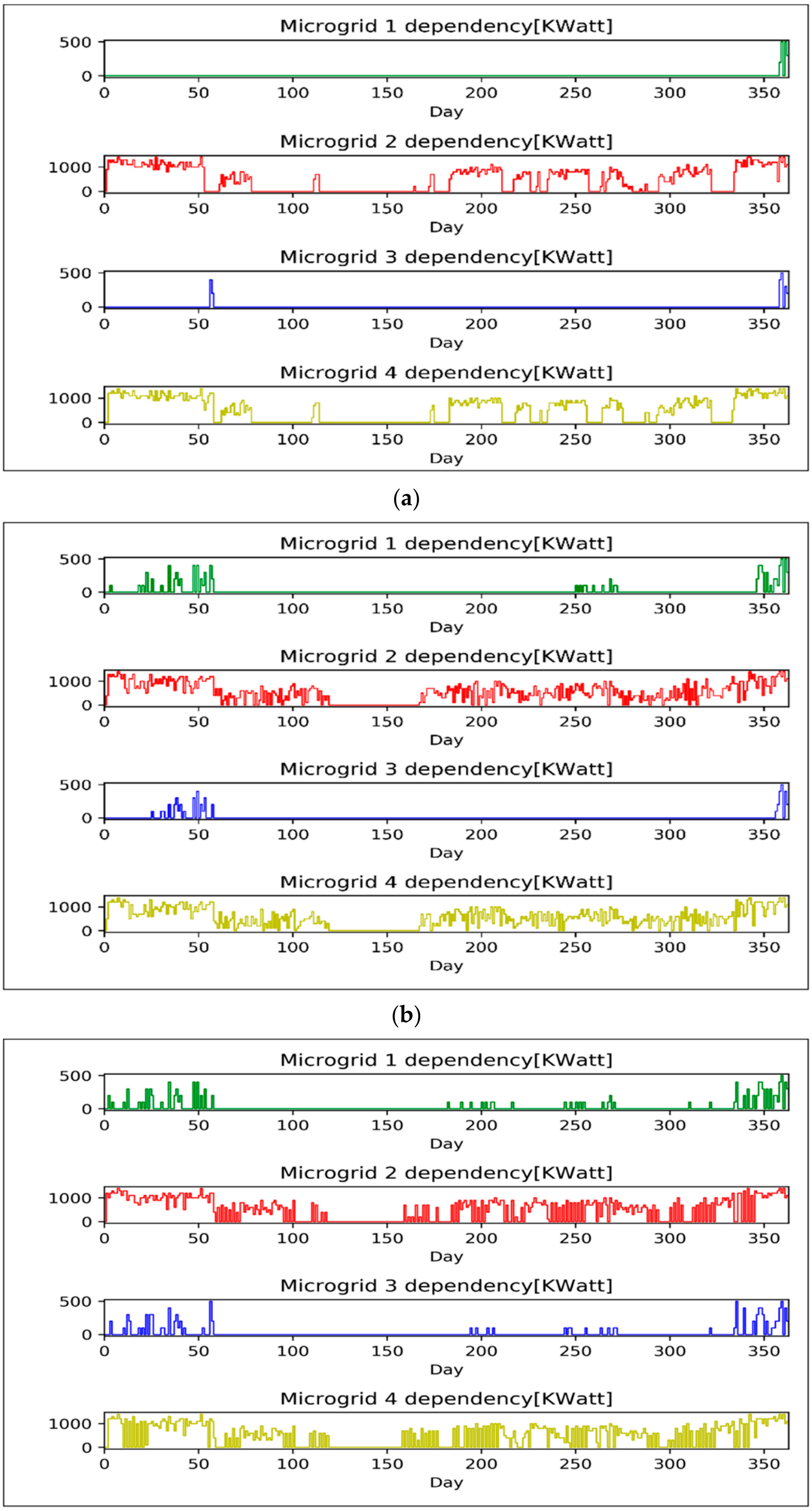


Figure 4. The DDQN learning curve showing score convergences for episode 2500 using the proposed double Kelly, EFD [15], and CCPV [15] strategies.

In Figure 5, the energy quantity of each MG received from the main grid is presented. In Figure 5a, it is confirmed that the proposed strategy can reach a significant reduction in the energy quantity from the main grid, which leads to a meaningfully lower dependency on the main grid. Therefore, there is a high possibility of making a self-sufficient system by the proposed strategy. It can be also interpreted that the double Kelly strategy based ETS acts intending to reduce the dependence of the entire MG by taking drastic trading actions in consideration of the future state while EFD and CCPV strategies only focus on the present state without looking at the long-term perspective. In Figure 5b,c, taken as a whole, the insufficient energy of the MG is supplied by the main grid and can be regarded as partially dependent on the main grid.

In Figure 6, the total quantity of dependent energy on the main grid is presented over a year. A lower value of total quantity represents lower dependency on the main grid. The independence performance of the proposed ETS can be improved by 17% compared to EFD-ETS and 7% compared to CCPV-ETS, respectively. Therefore, the aims of energy independence can be achieved by utilizing the proposed ETS scheme.

In Figure 7, total profits were presented over a year for the proposed ETS, EFD-ETS, and CCPV-ETS strategies. It is confirmed that the proposed ETS can achieve higher profits compared with EFD and CCPV strategies. In the proposed ETS, each MG competes with the others in different situations and aims at increasing social welfare by an equilibrium state. However, the EFD and CCPV strategies do not produce high profits because they are not competitive strategies.



(**a**)

(**b**)

(**c**)

Figure 5. Energy quantity received from the main grid by each MG in the optimized model; (**a**) proposed ETS (**b**) EFD-ETS (**c**) CCPV-ETS.

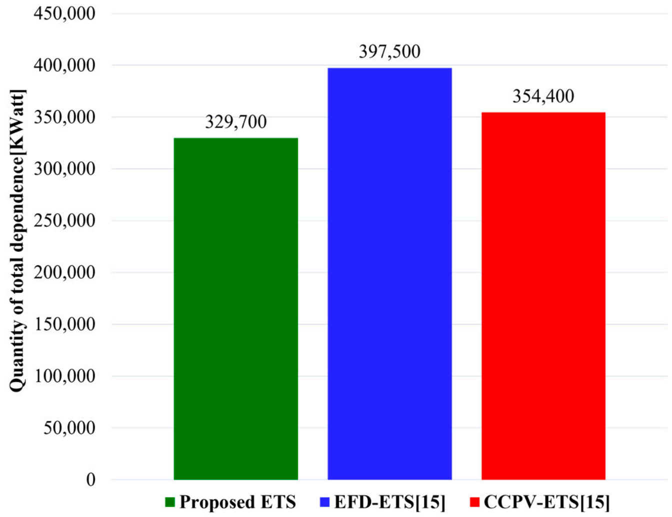


Figure 6. Quantity of total dependency of MGs on the main grid over a year by the proposed ETS, EFD-ETS, and CCPV-ETS.

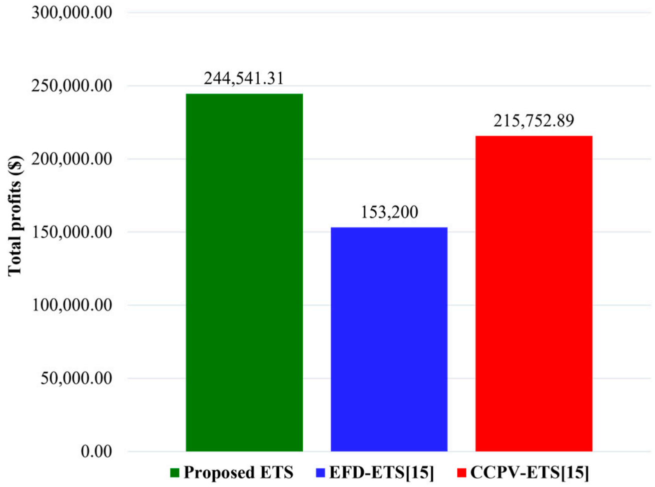


Figure 7. Total profits in MGs over a year for the proposed ETS, EFD-ETS, and CCPV-ETS.

4. Conclusions and Discussion

In this paper, the energy trading issue of establishing self-sufficient energy and an efficient MG system was described. The inefficiencies in energy trading systems of microgrids caused by uncertainties of non-stationary environments can be mitigated by the proposed double Kelly strategy. The DDQN algorithm-based ETS with a double Kelly strategy has been proposed in order to improve the profits and energy self-sufficiency in the microgrid's energy trading systems for the long-term. From the simulation results, it was verified that the proposed strategy can reduce the dependence on the main grid in non-stationary environments. Additionally, in terms of profitability, significant improvement was demonstrated via appropriate pricing and energy trading quantity. In conclusion, it is worth reiterating that the proposed ETS coordination, based on the day-ahead proper energy trading, can provide a promising approach to profitable and energy self-sufficient solutions for the future microgrid. The proposed scheme can affect establishing an efficient energy trading policy for alleviating the problems of energy imbalance.

A few limitations of the proposed scheme may include depreciation of the energy storage system with time and, additionally, the deprecation of the characteristics of applicable distributed energy resources to practical energy trading systems of the microgrid. It is expected that these limitations can be relieved or overcome by the following techniques:

efficient management of surplus energy for the energy storage systems, risk management of each energy resource, precise control of demand and supply system and more sophisticated prediction algorithms of energy trading environments.

Future research can be directed towards security of energy information, which could be improved by conducting distributed ETS rather than a centralized energy transaction control. Additionally, another study related to energy storage systems can be suggested for improving battery efficiency. In addition, different social factors and predictive algorithms can be considered for enhancing the performance of trading efficiency. The proposed energy trading scheme can find its application in establishing efficient and effective energy management and transaction systems for microgrid energy networks.

Author Contributions: Conceptualization, S.L.; Data curation, J.S., C.K. and S.K.; Methodology, S.L.; Project administration, J.K.; Software, S.L., J.S. and C.K.; Validation, Y.S.; Visualization, S.K.; Writing—original draft preparation, S.L.; Writing—review & editing, Y.S. and J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was, in part, supported by "Human Resources Program in Energy Technology (No. 20194010201830)" of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resources from the Ministry of Trade, Industry & Energy, Republic of Korea and in part supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1F1A1058266) and in part by Kwangwoon University in 2020.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shrivastwa, R.R.; Hably, A.; Melizi, K.; Bacha, S. Understanding microgrids and their future trends. In Proceedings of the 2019 IEEE International Conference on Industrial Technology (ICIT), Melbourne, Australia, 13–15 February 2019; pp. 1723–1728.
2. Shahidehpour, M.; Yan, M.; Shikhar, P.; Bahramirad, S.; Paaso, A. Blockchain for peer-to-peer transactive energy trading in networked microgrids: Providing an effective and decentralized strategy. *IEEE Electrif. Mag.* **2020**, *8*, 80–90. [CrossRef]
3. Meenual, T.; Usapein, P. Microgrid policies: A review of technologies and key drivers of Thailand. *Front. Energy Res.* **2021**, *9*, 48. [CrossRef]
4. Wang, W.; Lu, Z. Cyber security in the smart grid: Survey and challenges. *ScienceDirect* **2013**, *57*, 1344–1371. [CrossRef]
5. Hirsch, A.; Parag, Y.; Guerrero, J. Microgrids: A review of technologies, key drivers, and outstanding issues. *ScienceDirect* **2018**, *90*, 402–411. [CrossRef]
6. Zia, M.F.; Elbouchikhi, E.; Benbouzid, M.; Guerrero, J.M. Microgrid transactive energy systems: A perspective on design, technologies, and energy markets. In Proceedings of the IECON 2019—45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, 14–17 September 2019; pp. 5795–5800.
7. Cui, T.; Wang, Y.; Nazarian, S.; Pedram, M. An electricity trade model for microgrid communities in smart grid. In Proceedings of the 2014 IEEE Innovative Smart Grid Technologies (ISGT), Lumpur, Malaysia, 20–23 May 2014; pp. 1–5.
8. Colmenar-Santos, A.; De Palacio, C.; Enríquez-García, L.A.; López-Rey, Á. A methodology for assessing islanding of microgrids: Between utility dependence and off-grid systems. *Energies* **2015**, *8*, 4436–4454. [CrossRef]
9. Liu, Y.; Fang, Y.; Li, J. Interconnecting microgrids via the energy router with smart energy management. *Energies* **2017**, *10*, 1297. [CrossRef]
10. Zou, H.; Mao, S.; Wang, Y.; Zhang, F.; Chen, X.; Cheng, L. A survey of energy management in interconnected multi-microgrids. *IEEE Access* **2019**, *7*, 72158–72169. [CrossRef]
11. Bayramov, S.; Prokazov, I.; Kondrashev, S.; Kowalik, J. Household electricity generation as a way of energy independence of states—social context of energy management. *Energies* **2021**, *14*, 3407. [CrossRef]
12. Afzalan, M.; Jazizadeh, F. Quantification of demand-supply balancing capacity among prosumers and consumers: Community self-sufficiency assessment for energy trading. *Energies* **2021**, *14*, 4318. [CrossRef]
13. Kim, B.; Bae, S.; Kim, H. Optimal energy scheduling and transaction mechanism for multiple microgrids. *Energies* **2017**, *10*, 566. [CrossRef]

14. Chung, K.-H.; Hur, D. Towards the design of P2P energy trading scheme based on optimal energy scheduling for prosumers. *Energies* **2020**, *13*, 5177. [CrossRef]
15. Alabdullatif, A.M.; Gerding, E.H.; Perez-Diaz, A. Market design and trading strategies for community energy markets with storage and renewable supply. *Energies* **2020**, *13*, 972. [CrossRef]
16. Kazacos Winter, D.; Khatri, R.; Schmidt, M. Decentralized prosumer-centric P2P electricity market coordination with grid security. *Energies* **2021**, *14*, 4665. [CrossRef]
17. Levent, T.; Preux, P.; Le Pennec, E.; Badosa, J.; Henri, G.; Bonnassieux, Y. Energy management for microgrids: A reinforcement learning approach. In Proceedings of the 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), Bucharest, Romania, 29 September–2 October 2019; pp. 1–5.
18. Wang, H.; Huang, T.; Liao, X.; Abu-Rub, H.; Chen, G. Reinforcement learning in energy trading game among smart microgrids. *IEEE Trans. Ind. Electron.* **2016**, *63*, 5109–5119. [CrossRef]
19. Lu, X.; Xiao, X.; Xiao, L.; Dai, C.; Peng, M.; Poor, H.V. Reinforcement learning-based microgrid energy trading with a reduced power plant schedule. *IEEE Internet Things J.* **2019**, *6*, 10728–10737. [CrossRef]
20. Ji, Y.; Wang, J.; Xu, J.; Fang, X.; Zhang, H. Real-time energy management of a microgrid using deep reinforcement learning. *Energies* **2019**, *12*, 2291. [CrossRef]
21. Bell, V.A.; Kay, A.L.; Jones, R.G.; Moore, R.J. Development of a high resolution grid-based river flow model for use with regional climate model output. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 532–549. [CrossRef]
22. Kelly, J.L. A new interpretation of information rate. *Bell Syst. Tech. J.* **1956**, *35*, 917–926. [CrossRef]
23. Yoon, T.; Jo, S. Analysis on the seasonal patterns of electricity demand for housing and its implications. *Korea Energy Econ. Inst.* **2016**, *13*, 130–132.
24. Chris, W. Learning from delayed rewards. In *Doctoral Dissertation*; King's College: London, UK, 1989.
25. Prabuchandran, K.J.; Meena, S.K.; Bhatnagar, S. Q-learning based energy management policies for a single sensor node with finite buffer. *IEEE Wirel. Commun. Lett.* **2013**, *2*, 82–85. [CrossRef]
26. Watkins, C.J.C.H.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]
27. Volodymyr, M.; Koray, K.; David, S.; Alex, G.; Ioannnis, A.; Daan, W.; Martin, R. Playing Atari with deep reinforcement learning. *arXiv* **2013**, arXiv:1312.5602v1.
28. Hado, V.H.; Arthur, G.; David, S. Deep reinforcement learning with double Q-learning. *arXiv* **2015**, arXiv:1509.06461v3.
29. Hado, V.H. Double Q-learning. *Adv. Neural Inf. Process. Syst.* **2010**, *23*, 2613–2621.
30. Statistics Korea. Available online: http://kostat.go.kr/portal/eng/index.action (accessed on 10 August 2021).
31. Mingyeong, K. Factors of changing household power consumption in Seoul and reduction plan. *Seoul Inst. Policy Rep.* **2013**, *149*, 9–11.
32. Soysal, O.A.; Soysal, H.S. From wind-solar energy educational demo system (WISE) to sustainable energy research facility (SERF). In Proceedings of the IEEE Power & Energy Society General Meeting (PES), Calgary, AB, Canada, 26–30 July 2009.
33. Derek, D.W.; Toshiyuki, S. Assessment of large commercial rooftop photovoltaic system installations: Evidence from California. *Appl. Energy* **2017**, *188*, 45–55.
34. Actual Operating Plant Power Generation. Available online: https://www.mal-eum.com/sample (accessed on 21 April 2021).
35. Matthew, L.; David, A.; Miaomiao, H. Seasonal variation in household electricity demand: A comparison of monitored and synthetic daily load profiles. *Energy Build.* **2018**, *179*, 292–300.
36. Harish, V.S.K.V.; Anwer, N.; Kumar, A. Development of a peer to peer electricity exchange model in microgrids for rural electrification. In Proceedings of the 2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC), Greater Noida, India, 18–19 October 2019; pp. 259–263.