energies

MDPI

Article

Q-Learning-Based Operation Strategy for Community Battery Energy Storage System (CBESS) in Microgrid System

Van-Hai Bui, Akhtar Hussain and Hak-Man Kim *

Department of Electrical Engineering, Incheon National University, 12-1 Songdo-dong, Yeonsu-gu, Incheon 22012, Korea; buivanhaibk@inu.ac.kr (V.-H.B.); hussainakhtar@inu.ac.kr (A.H.)

* Correspondence: hmkim@inu.ac.kr; Tel.: +82-32-835-8769; Fax: +82-32-835-0773

Received: 11 April 2019; Accepted: 8 May 2019; Published: 10 May 2019

Abstract: Energy management systems (EMSs) of microgrids (MGs) can be broadly categorized as centralized or decentralized EMSs. The centralized approach may not be suitable for a system having several entities that have their own operation objectives. On the other hand, the use of the decentralized approach leads to an increase in the operation cost due to local optimization. In this paper, both centralized and decentralized approaches are combined for managing the operation of a distributed system, which is comprised of an MG and a community battery storage system (CBESS). The MG is formed by grouping all entities having the same operation objective and is operated under a centralized controller, i.e., a microgrid EMS (MG-EMS). The CBESS is operated by using its local controller with different operation objectives. A Q-learning-based operation strategy is proposed for optimal operation of CBESS in both grid-connected and islanded modes. The objective of CBESS in the grid-connected mode is to maximize its profit while the objective of CBESS in islanded mode is to minimize the load shedding amount in the entire system by cooperating with the MG. A comparison between the Q-learning-based strategy and a conventional centralized-based strategy is presented to show the effectiveness of the proposed strategy. In addition, an adjusted epsilon is also introduced for epsilon-greedy policy to reduce the learning time and improve the operation results.

Keywords: artificial intelligence; battery energy storage system; energy management system; microgrid operation; optimization; Q-learning-based operation

1. Introduction

Microgrid (MG) is a small-scale electric power system, which can be operated both in islanded and grid-connected modes. The operation of the MG is generally carried out by an energy management system (EMS) [1,2]. In recent years, the development of centralized EMSs has been extensively studied and used for the operation of MGs [3–5]. However, these centralized EMSs are facing many problems such as computational burden and complexity in communication networks especially, when the numbers of control devices increase. Therefore, this may lead to scalability issues for future expansion [6]. In addition, there are several entities in the MG system that have different owners and different operation objectives [7]. Therefore, it is difficult to provide a common operation objective for the operation of the entire system. Recently, decentralized EMSs are becoming popular due to their ability to overcome the limitations of centralized EMSs [8–10]. In decentralized EMSs, each entity in the system is monitored and controlled by a local controller, which only communicates with its neighboring controllers via a communication network. Due to the lack of detailed information about the other entities in the system, the solution may not be globally optimal [11].

Therefore, the use of centralized or decentralized EMSs is not efficient for the operation of a distributed system with different entities having diverse operation objectives. A potential solution could be developing an EMS with a combination of both of these approaches to take advantage of each framework. In this approach, a group of entities having the same operation objectives are operated under a centralized controller, while the other entities having different operation objectives are operated by using other local controllers. Centralized controllers gather all system information and determine the optimal schedule for each component by using mathematical programming. On the other hand, the reinforcement learning (RL) approach has been introduced for the distributed operation of independent entities having local operation objectives in case the independent entities do not have complete information of the environment [12–17]. In RL, agents learn to achieve a given task by interacting with their environment. Since the agents do not require any model of the environment, they only need to know the existing states and possible actions in each state. This method drives the learning process based on penalties or rewards assessed on a sequence of actions taken in response to the environment dynamics [17,18]. In contrast to the conventional distributed methods, learning-based methods can be easily adapted with a real-time problem after the off-line training process. In RL, Q-learning is a popular method and is widely used for the optimal operation of microgrids [19–23]. A fitted Q-iteration-based algorithm has been proposed in [19] for a BESS. A data-driven method is utilized in [19] and it uses a state-action value function to optimize a scheduling plan for the BESS in grid-connected mode. An RL-based energy management algorithm has been proposed by [20] to reduce the operation cost of a smart energy building under unknown future information. The authors in [21] have proposed a multiagent RL-based distributed optimization of a solar MG by optimizing the schedule of an energy storage system. A two steps-ahead RL algorithm has been developed in [22] to plan battery scheduling. By using this method, the utilization rate of the battery is increased during high electricity demand while the utilization rate of the wind turbine for local demand is also increased to reduce the consumer dependence on the utility grid. The authors in [23] have presented an improved RL method to minimize the operation cost of an MG in the grid-connected mode.

However, most of the existing Q-learning-based operation methods have been developed for optimal operation of an agent in the grid-connected mode only for a particular objective, i.e., maximization of profit (competitive model). However, in the case of islanded mode, they may have adverse effects and reduce the reliability of the entire system, such as the increased load shedding amount. Therefore, an energy management strategy, which is applicable for both grid-connected and islanded modes with different objectives need to be developed. In addition, most of the existing literature on Q-learning-based methods have been developed for optimal operation of a single MG and only focused on the operation of the local components of an MG. Adjacent MGs can be interconnected to form a multi-microgrid system to improve network reliability by sharing power among MGs and other community entities [11]. However, the power transfers between other community entities and among MGs of the network have not been considered in the existing Q-learning-based operation methods [19–23]. Therefore, the existing methods are not suitable to apply for multi-microgrid systems.

In order to overcome the problems mentioned above, a Q-learning-based energy management strategy is developed in this paper for managing the operation of a distributed system. The system is comprised of an MG and a community BESS (CBESS). A microgrid EMS (MG-EMS) is used for managing the operation of the MG while a Q-learning-based operation strategy is proposed for the optimal operation of the CBESS. In contrast to the existing literature [19–23], where only grid-connected mode operation is considered, both grid-connected and islanded mode operations are considered in this study. The objective in grid-connected mode is to maximize the profit of the CBESS via optimal charging/discharging decisions by trading power with the utility grid and other MGs of the network. However, in islanded mode, the objective is to minimize the load shedding amount in the network by cooperating with the MGs of the network. Due to the consideration of power trading among community resources and MGs of the network, the proposed method can be easily extended for multi-microgrid systems. However, the existing methods in the literature [19–23] focus on simplified single MGs, cannot be applied for networked MGs. To analyze the effectiveness of the proposed Q-learning-based optimization method, the operation results of the proposed method are compared with the conventional centralized EMS results. Simulation results have proved that the proposed method can get similar results with the centralized EMS results, despite being a decentralized approach. Finally, an adjusted epsilon method is applied in the epsilon-greedy policy to reduce the learning time and improve the operation results.

2. System Model

2.1. Test System Configuration

Figure 1 describes a test system configuration, which is comprised of an MG and a CBESS. In this study, MG is a group of entities having the same operation objectives, such as CDG, RDGs, BESS, and loads, which are operated under a centralized controller (i.e., MG-EMS), while the CBESS having different operation objectives is operated by using its local controller. A Q-learning-based operation strategy is proposed for CBESS.

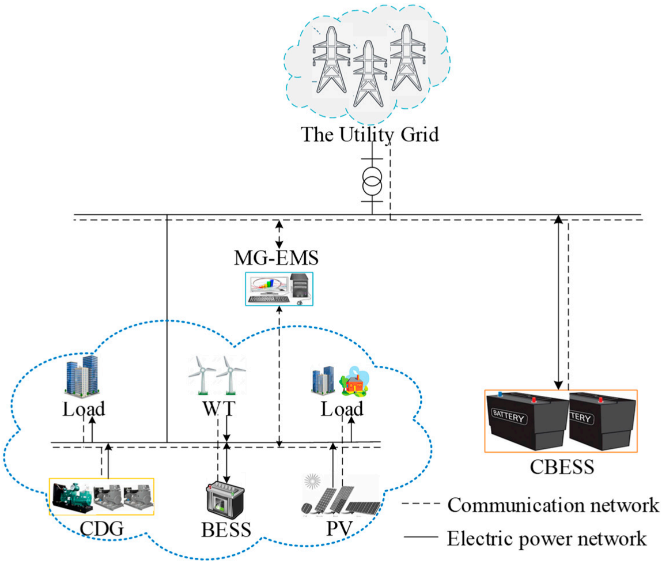


Figure 1. Test system configuration.

In the grid-connected mode, power can be traded among MG, CBESS, and the utility grid for minimizing the total operation cost. In islanded mode, the MG system is not connected to the utility grid, CBESS and MG can trade power to minimize load shedding in the network. The MG considered in this study consists of a controllable distributed generator (CDG), a renewable distributed generator (RDG) system, a BESS as the energy storage device, and residential loads. MG is operated by an MG-EMS for minimizing its operation cost. In grid-connected mode, MG-EMS communicates with the utility grid to get the buying/selling price and decides the amount of buying/selling power to be traded with the utility grid and CBESS. In islanded mode, MG cannot trade with the utility grid. Thus MG cooperatively operates with CBESS to minimize the load shedding amount. The detailed algorithms for both operation modes are explained in the following section.

2.2. Q-Learning-Based Operation Strategy for CBESS

Q-learning is a model-free reinforcement learning where an agent explores the environment and finds the optimal way to maximize the cumulative reward [19–23]. In Q-learning, the agent does not need to have any model of the environment. It only needs to know the existing states and possible actions in each state. Each state-action pair is assigned an estimated value, called a Q value, which is the brain of the agent. A Q-table represents all the knowledge of an agent about the environment. When the agent comes to a state and takes an action, it receives a reward. The reward is used to update the Q value of the agent. The overall Q-learning principle diagram for CBESS is summarized in Figure 2.

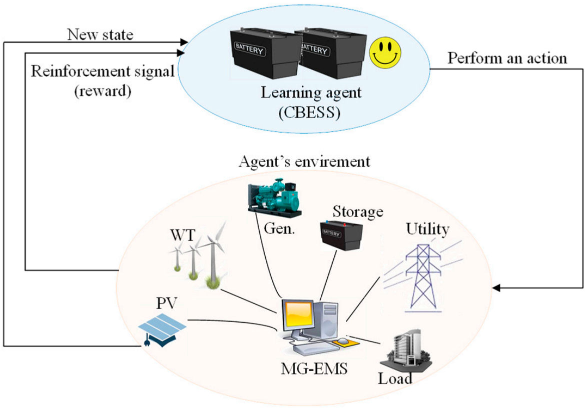


Figure 2. Q-learning principle diagram.

Figure 3 shows the states and possible actions of CBESS. In this study, each state is a vector *s* with three features: time interval, SOC of CBESS, and market price signal at the current interval. The CBESS agent starts from an initial state with an initial value of SoC and initial interval *t*. The CBESS is operated for a 24-hour scheduling horizon and each time interval is set to be one hour. Therefore, the initial interval *t* is usually taken as the first interval (t = 1). CBESS chooses a random action and receives a reward according to the action. CBESS will perform several actions until reaching the goal state. CBESS can be either in charging, discharging, or idle mode in each state. Thus, SoC of CBESS can also be increased, decreased, or kept the same with the previous state depending on the choosing action and charging/discharging amount. This amount could be any values in the operation bounds of CBESS. There are some special cases, as following.

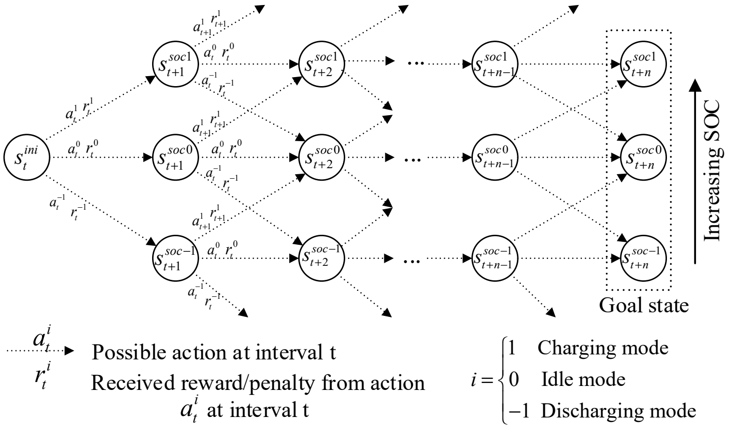


Figure 3. Possible states and actions of a community battery storage system (CBESS).

If the CBESS is fully charged, it cannot charge more. In case the CBESS decides to charge, it is facing a high penalty to avoid this action in the future and the suitable actions are discharging or idle mode. In contrast, if the CBESS is fully discharged, the CBESS cannot discharge more. It is facing a high penalty for a discharge decision and the suitable actions are charging or idle mode. The objective of CBESS is to maximize its profit by optimal charging/discharging decisions. This can be obtained by maximizing the cumulative reward following the Q-table. The reward function for CBESS is determined by Algorithm 1 based on the chosen action. In grid-connected mode, the charging/discharging price signals are taken from the market price signals. However, in islanded mode, MG-EMS decides the charging/discharging price signals to increase the utilization of CBESS for minimizing load shedding amount. For instance, during off-peak load intervals, the charging price is low, CBESS buys surplus power for charging mode. During the peak load intervals, in order to avoid load shedding in the MG system, the discharging price is high. Thus CBESS discharges power to fulfill the shortage power.

Algorithm 1: Reward Function for CBESS

```
Input: a state $s$ and action $a$
if $a$ = "charge" do
    if $SoC = SoC_{max}$ do
        $r$ = a high penalty
    else
        $r = -P_{char} \cdot price$
    end if
else if $a$ = "idle" do
    $r = 0$
else $a$ = "discharge" do
    if $SoC = SoC_{max}$ do
        $r$ = a high penalty
    else
        $r = P_{dis} \cdot price$
    end if
end if
```

The learning strategy for CBESS is summarized in Algorithm 2. Firstly, the discount factor (γ) and learning rate (α) are taken as input data for the algorithm. The value of the discounted factor contributes to determining the value of the future reward. This value varies from 0 to 1, in case the value is near 0, the immediate reward is given preference and in case the value is near 1, the importance of future rewards is increased. The learning rate affects the speed of convergence of Q values. The value of the learning rate also varies from 0 to 1. However, it should be a small value to ensure the convergence of the model. Therefore, the value of the discount factor and learning rate are 0.99 and 0.1, respectively. The CBESS agent explores the environment for a large number of episodes. In each episode, the CBESS agent starts from an initial state (interval $t = 1$ with an initial value of SoC), the agent performs several actions until reaching the goal state and updates its knowledge (Q-table). The choosing action is based on the epsilon-greedy policy, which is a way of selecting random actions with uniform distribution from a set of possible actions [24,25].

Algorithm 2: Q-Learning-Based Operation Strategy for CBESS

```
Input data: setting $\alpha, \gamma$
Initialize a $Q$-table arbitrarily ·
for $episode < episode_{max}$ do
    while $s$ is not terminal do
        Initialize a starting state $s$ (i.e., interval = 1, $SoC = SoC_{ini}$, and market price ($interval = 1$))
        Select a possible action $a$ from $s$ using $\varepsilon$-greedy policy
        Take action $a$ and observe reward $r$, and come to state $s'$ (Algorithm 1)
        if $s'$ is not in Q-table do
            Initialize $Q(s', a_i) = 0$
        end if
        Update the Q-table: $Q(s,a) \leftarrow Q(s,a) + \alpha \cdot [r + \gamma \cdot \max_{a'} Q(s',a') - Q(s,a)]$
        Update state $s$ to the next state $s'$ with new $SoC$
        if $interval = 24$ do
            $s$ is terminal
        end if
    end while
end for
```

Using this policy, the agent selects a random action with ε probability and an action with a probability of $(1-\varepsilon)$ that gives a maximum reward in a given state. After performing an action, the Q value is updated by using Equation (1).

$$\begin{gathered} Q(s,a) \leftarrow Q(s,a) + \alpha.[r + \gamma.\max_{a\prime} Q(s\prime, a\prime) - Q(s,a)] \\ s \leftarrow s\prime \end{gathered} \tag{1}$$

The current state is moved to a next state with updated SoC. The process for each episode is terminated when the goal state is reached. The CBESS can find optimal actions after exploring the environment with a large number of episodes.

2.3. Operation Strategy for Microgrid and CBESS

Figure 4 shows the detailed operation strategy for MG and CBESS. In the grid-connected mode, MG-EMS receives the market price signals from the utility grid. MG-EMS also gathers all information of the MG system and performs optimization to minimize the total operation cost. The amount of surplus/shortage power is determined based on the optimal results. Then the MG-EMS waits for information from the other external systems. The CBESS also learns from the environment and updates its knowledge according to Algorithm 2. The amount of charging/discharging power is determined at the end of the process. All information for trading amount with the external system is informed by CBESS. After gathering the information from the CBESS and the utility grid, MG-EMS decides the amount of buying/selling power from/to the utility grid and CBESS and informs the optimal results to its components. In islanded mode, there is no connection to the utility grid. Load shedding could be implemented to maintain the power balance. In order to reduce the amount of load shedding, CBESS could be in cooperative operation mode with the MG. After performing optimization by MG-EMS, the information of surplus/shortage power is determined in each interval of time. Similarly, CBESS learns with a large number of episodes for optimizing its operation based on the feedback from MG-EMS. The final operation of CBESS is informed to MG-EMS with the charging/discharging amount. Finally, MG-EMS reschedules the operation of all the components based on the charging/discharging amount from CBESS. Load shedding is implemented for maintaining the power balance in the whole system in case of having a shortage of power.

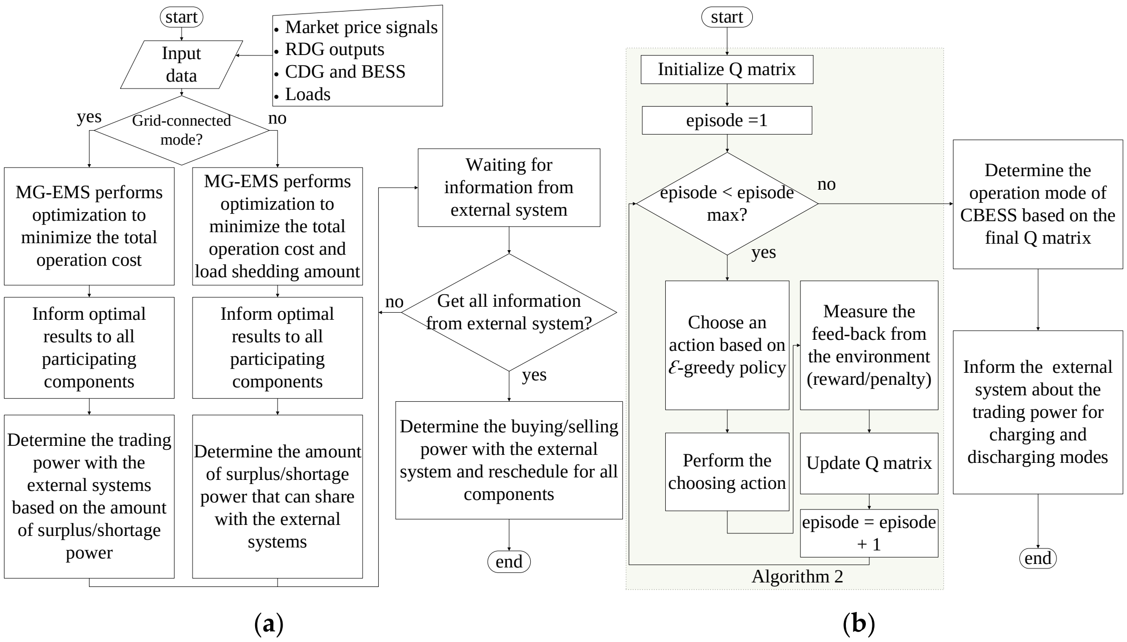

start
Input data
• Market price signals
• RDG outputs
• CDG and BESS
• Loads
Grid-connected mode?
yes
no
MG-EMS performs optimization to minimize the total operation cost
MG-EMS performs optimization to minimize the total operation cost and load shedding amount
Inform optimal results to all participating components
Inform optimal results to all participating components
Determine the trading power with the external systems based on the amount of surplus/shortage power
Determine the amount of surplus/shortage power that can share with the external systems
Waiting for information from external system
Get all information from external system?
no
yes
Determine the buying/selling power with the external system and reschedule for all components
end
(a)
start
Initialize Q matrix
episode =1
episode < episode max?
no
yes
Choose an action based on ℰ-greedy policy
Measure the feed-back from the environment (reward/penalty)
Perform the choosing action
Update Q matrix
episode = episode + 1
Algorithm 2
Determine the operation mode of CBESS based on the final Q matrix
Inform the external system about the trading power for charging and discharging modes
end
(b)

Figure 4. Flowchart for operation strategy: (**a**) MG-EMS; (**b**) CBESS.

2.4. Mathematical Model

In this section, a mixed integer linear program (MILP)-based formulation is presented for day-ahead scheduling (i.e., T = 24 h) for all components in MG system for both grid-connected and islanded modes. In grid-connected mode, the objective function (2) is to minimize the total operation cost associated with the fuel cost, start-up/shut-down cost of CDGs, and cost/benefit of purchasing/selling power from/to the utility grid, as shown in Equation (2).

$$\min\left\{\begin{array}{c}\sum\limits_{i\in I}\sum\limits_{t\in T}\left(C_i^{CDG}\cdot P_{i,t}^{CDG}+y_{i,t}\cdot C_i^{SU}+z_{i,t}\cdot C_i^{SD}\right)\\+\sum\limits_{t\in T}\left(PR_t^{Buy}\cdot P_t^{Buy}\right)-\sum\limits_{t\in T}\left(PR_t^{Sell}\cdot P_t^{Sell}\right)\end{array}\right\}\tag{2}$$

The constraints associated with CDGs include Equations (3)–(8). Constraint (3) enforces the upper and lower operation bounds of CDGs. Equation (4) gives the on/off status of CDGs. The start-up and shut-down modes are determined by using constraints (5) and (6) based on the on/off status of CDGs. The bounds for ramp up/ramp down rates of CDGs are enforced by Equations (7) and (8), respectively.

$$u_{i,t}\cdot P_i^{\min}\le P_{i,t}^{CDG}\le u_{i,t}\cdot P_i^{\max}\qquad\forall i\in I,t\in T\tag{3}$$

$$u_{i,t}=\left\{\begin{array}{ll}1 & CDG\ is\ on\\0 & CDG\ is\ off\end{array}\right.\qquad\forall i\in I,t\in T\tag{4}$$

$$y_{i,t}=\max\left\{(u_{i,t}-u_{i,t-1}),0\right\}\qquad\forall i\in I,t\in T\tag{5}$$

$$z_{i,t}=\max\left\{(u_{i,t-1}-u_{i,t}),0\right\}\qquad\forall i\in I,t\in T\tag{6}$$

$$P_{i,t}^{CDG}-P_{i,t-1}^{CDG}\le RU_i\cdot(1-y_{i,t})+P_i^{\min}\cdot y_{i,t}\quad\forall i\in I,t\in T\tag{7}$$

$$P_{i,t-1}^{CDG}-P_{i,t}^{CDG}\le RD_i\cdot(1-z_{i,t})+P_i^{\min}\cdot z_{i,t}\quad\forall i\in I,t\in T\tag{8}$$

The power balance between the power sources and power demand is given by Equation (9). The buying/selling power is the amount of power trading with the external systems, which is divided into trading with the utility grid or CBESS, as given in Equations (10) and (11), respectively.

$$P_t^{PV}+P_t^{WT}+\sum_{i\in I}P_{i,t}^{CDG}+P_t^{Buy}+P_t^{B-}=\sum_{l\in L}P_{l,t}^{Load}+P_t^{Sell}+P_t^{B+}\quad\forall t\in T\tag{9}$$

$$P_t^{Sell}=P_t^{Sell\_Grid}+P_t^{Sell\_CBESS}\qquad\forall t\in T\tag{10}$$

$$P_t^{Buy}=P_t^{Buy\_Grid}+P_t^{Buy\_CBESS}\qquad\forall t\in T\tag{11}$$

The constraints related to BESS include Equations (12)–(16). Constraint (12) and (13) are the maximum charging/discharging power of the BESS. The value of SoC is updated by Equation (14) after charging/discharging power at each interval of time. Equation (15) shows the value of SoC is set by initial SoC at the first interval of time (t = 1). The operation bounds of BESS are enforced by (16).

$$0\le P_t^{B+}\le P_B^{Cap}\cdot\left(SoC_{\max}^B-SoC_{t-1}^B\right)\cdot\frac{1}{1-L^{B+}}\quad\forall t\in T\tag{12}$$

$$0\le P_t^{B-}\le P_B^{Cap}\cdot\left(SoC_{t-1}^B-SoC_{\min}^B\right)\cdot(1-L^{B-})\quad\forall t\in T\tag{13}$$

$$SoC_t^B=SoC_{t-1}^B-\frac{1}{P_B^{Cap}}\cdot\left(\frac{1}{1-L^{B-}}\cdot P_t^{B-}-P_t^{B+}\cdot(1-L^{B+})\right)\quad\forall t\in T\tag{14}$$

$$SoC_{t-1}^B=SoC_{ini}^B\qquad\text{if } t=1\tag{15}$$

$$SoC_{\min}^{B} \leq SoC_{t}^{B} \leq SoC_{\max}^{B} \qquad \forall t \in T \tag{16}$$

In grid-connected mode, CBESS also optimizes its operation to maximize its profit. The CBESS is decided to charge power from the utility grid during off-peak price intervals or from the MG. It is in discharging mode during peak price intervals. The constraints for CBESS in grid-connected mode are shown in Equations (17)–(23). The total CBESS charging/discharging amount are the sum of charging/discharging power from/to both the utility grid and MG, as shown in Equations (17) and (18). The charging and discharging bounds are given by Equations (19) and (20). In this paper, the maximum charging/discharging power is 10% of the capacity of CBESS at each interval of time [22]. The value of SoC of CBESS is updated by using Equations (21) and (22). Finally, the operation bounds of CBESS is enforced by Equation (23).

$$P_{Grid,t}^{CB+} + P_{MG,t}^{CB+} = P_{t}^{CB+} \qquad \forall t \in T \tag{17}$$

$$P_{Grid,t}^{CB-} + P_{MG,t}^{CB-} = P_{t}^{CB-} \qquad \forall t \in T \tag{18}$$

$$0 \leq P_{t}^{CB+} \leq \min\left\{0.1 \cdot P_{CB}^{Cap}, P_{CB}^{Cap} \cdot \left(SoC_{\max}^{CB} - SoC_{t-1}^{CB}\right) \cdot \frac{1}{1 - L^{CB+}}\right\} \forall t \in T \tag{19}$$

$$0 \leq P_{t}^{CB-} \leq \min\left\{0.1 \cdot P_{CB}^{Cap}, P_{CB}^{Cap} \cdot \left(SoC_{t-1}^{CB} - SoC_{\min}^{CB}\right) \cdot (1 - L^{CB-})\right\} \forall t \in T \tag{20}$$

$$SoC_{t}^{CB} = SoC_{t-1}^{CB} - \frac{1}{P_{CB}^{Cap}} \cdot \left(\frac{1}{1 - L^{CB-}} \cdot P_{t}^{CB-} - P_{t}^{CB+} \cdot (1 - L^{CB+})\right) \forall t \in T \tag{21}$$

$$SoC_{t-1}^{CB} = SoC_{ini}^{CB} \qquad \text{if } t = 1 \tag{22}$$

$$SoC_{\min}^{CB} \leq SoC_{t}^{CB} \leq SoC_{\max}^{CB} \qquad \forall t \in T \tag{23}$$

In islanded mode, the system is disconnected from the utility grid. MG system can only trade its surplus/shortage power with CBESS. In peak intervals, MG and CBESS could not fulfill the power demand in the system. Therefore, the load shedding should be performed to keep the power balance in the system. In order to reduce the load shedding amount, MG-EMS performs optimization for minimizing both the total operation cost and the load shedding amount. The cost objective function is changed to (24) with the generation cost and the penalty for load shedding. The power balance of the power source and power demand is given by Equation (25) for the islanded mode. Additionally, the objective function (24) is also constrained by Equations (3)–(8) and Equations (12)–(16).

$$\min\left\{\begin{array}{c} \sum\limits_{i \in I} \sum\limits_{t \in T} \left(C_{i}^{CDG} \cdot P_{i,t}^{CDG} + y_{i,t} \cdot C_{i}^{SU} + z_{i,t} \cdot C_{i}^{SD}\right) \\ + \sum\limits_{t \in T} \left(C_{t}^{pen} \cdot P_{t}^{Short}\right) - \sum\limits_{t \in T} \left(C_{t}^{trade} \cdot P_{t}^{Sur}\right) \end{array}\right\} \tag{24}$$

$$P_{t}^{PV} + P_{t}^{WT} + \sum_{i \in I} P_{i,t}^{DG} + P_{t}^{B-} = \sum_{l \in L} P_{l,t}^{Load} + P_{t}^{B+} + P_{t}^{Sur} - P_{t}^{Short} \qquad \forall t \in T \tag{25}$$

In the islanded mode, the objective of CBESS is to reduce the amount of shortage power in MG system by optimal charging/discharging mode decisions. The CBESS is decided to charge surplus power from the MG system and discharge during intervals having shortage power. Constraints (26) and (27) show the bounds for charging/ discharging amount at an interval of time. These constraints also ensure that the charging mode is possible when the MG has surplus power, while the discharging mode is possible when the MG has shortage power. Additionally, the CBESS are also constrained by Equations (21)–(23) for updating the value of SoC and operation bounds of CBESS.

$$0 \leq P_{t}^{CB+} \leq \min\left\{0.1.P_{CB}^{Cap}, P_{CB}^{Cap} \cdot \left(SoC_{\max}^{CB} - SoC_{t-1}^{CB}\right) \cdot \frac{1}{1 - L^{CB+}}, P_{t}^{Sur}\right\} \forall t \in T \tag{26}$$

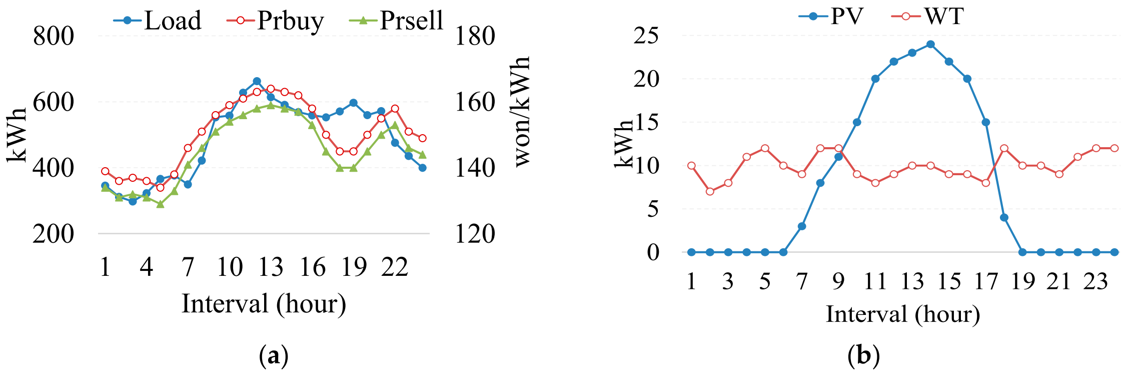
$$0 \leq P_{t}^{CB-} \leq \min\left\{0.1.P_{CB}^{Cap}, P_{CB}^{Cap} \cdot \left(SoC_{t-1}^{CB} - SoC_{\min}^{CB}\right) \cdot (1 - L^{CB-}), P_{t}^{Short}\right\} \forall t \in T \tag{27}$$

In this paper, a Q-learning-based operation strategy for CBESS is proposed for the optimal operation of CBESS. To show the effectiveness of the Q-learning-based operation, the results of Q-learning-based operation methods are compared with the results of the centralized operation method. The detailed numerical results are presented in the following section.

3. Numerical Results

3.1. Input Data

In this study, the test MG system has a PV, a WT, a CDG, a BESS, and load demand, as shown in Figure 1. The MG is interconnected with a CBESS and the utility grid. The system can be operated in both grid-connected and islanded modes. The analysis is conducted for a 24-hour scheduling horizon (T = 24 h) and each time interval is set to be 1 hour. The MILP-based model for MG is implemented in Python integrated with CPLEX 12.6 [26]. The Q-learning-based model for CBESS is also implemented in Python. The market price signals, load profile, and the total output of RDGs are shown in Figure 5a,b, respectively. The information of the CDG unit, BESS, and CBESS are tabulated in Table 1. The operation bounds of BESS and CBESS were chosen as [0%, 100%], same as [27]. The detailed numerical results are shown in the following sections for both grid-connected and islanded modes.



(**a**) (**b**)

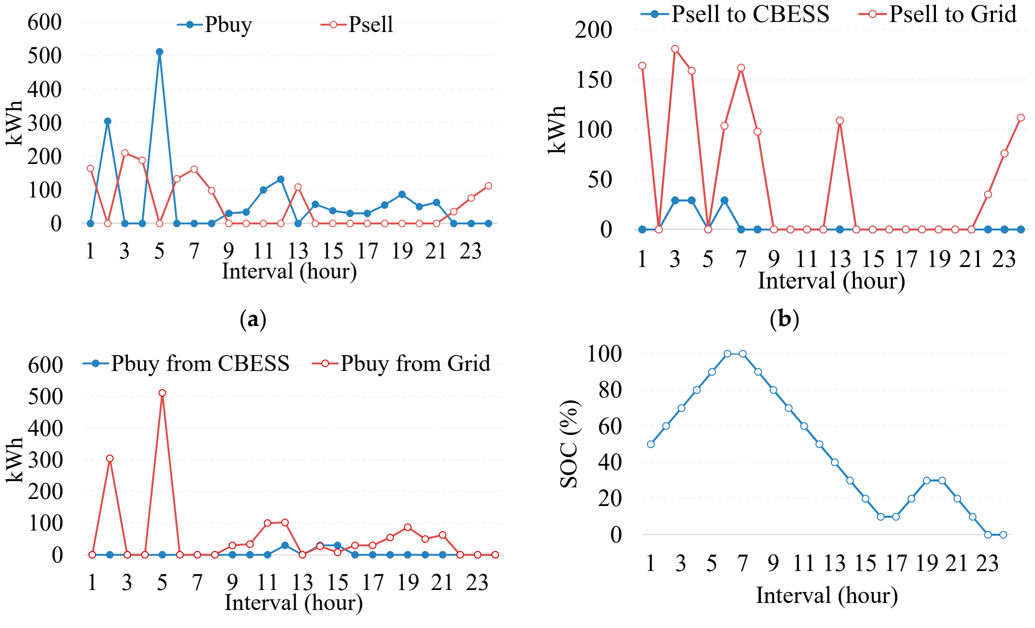
Figure 5. Input data: (**a**) market price signals and load profile; (**b**) output power of the renewable distributed generator (RDG).

Table 1. The detail parameters of BESS, CBESS, and controllable distributed generator (CDG).

Parameters	BESS	CBESS	Parameters	CDG
Max. P^{Cap} (kWh)	200	300	Max. $P^{\max}$(kWh)	500
Initial $P^{Cap} \cdot SoC_{ini}$ (kWh)	50	150	Min. $P^{\min}$ (kWh)	0
Min. $P^{Cap} \cdot SoC_{\min}$ (kWh)	0	0	Cost/kWh C^{CDG} (KRW)	136
Char. Loss L^{+} (%)	5	3	Start-up cost C^{SU} (KRW)	200
Dis. Loss L^{-} (%)	5	3	Shut-down cost C^{SD} (KRW)	100

3.2. Operation of the System in Grid-Connected Mode

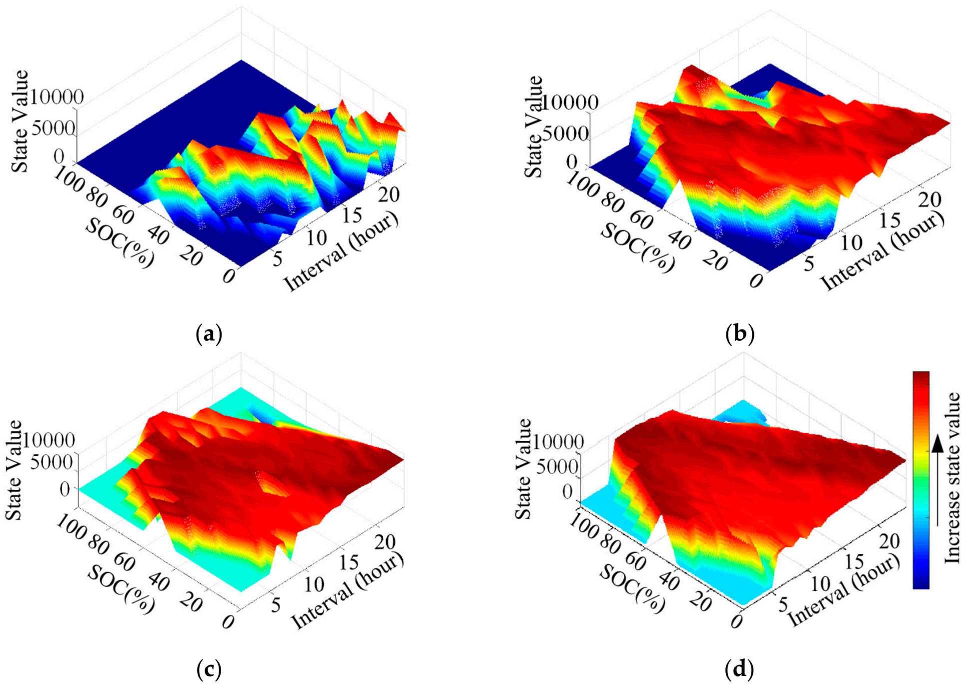
This section presents the operation of the MG and CBESS in grid-connected mode. The MG-EMS performs optimization to minimize the total operation cost of the MG. The amount of buying/selling power is determined based on the amount of surplus/shortage power in the MG system, as shown in Figure 6a. The buying/selling power of the MG is traded with two external systems, i.e., CBESS and the utility grid. It can be observed from Figure 6b that the CBESS always decides to import power from cheaper resources. During intervals 3, 4, 6, the generation cost of CDG is less than the buying prices from the utility grid. Therefore, CBESS decides to charge surplus power from MG instead of buying from the utility grid. Figure 6c shows the buying power of the MG system. The MG decides to import power from the external systems for minimizing the total operation cost. At intervals 2 and 5, MG imports power from the utility grid to fulfill load amount with cheaper price compared with the generation cost. During peak price intervals (12–15), MG also imports power from CBESS to fulfill the shortage power and reduce the amount of buying power from the utility grid. The optimal operation of CBESS is shown in Figure 6d by using the centralized-based approach. The Q-learning-based operation of CBESS is compared with the centralized-based operation to show the effectiveness of the proposed method.



(**a**) (**b**) (**c**) (**d**)

Figure 6. External trading from the microgrid (MG): (**a**) total trading amount from the MG; (**b**) selling power from the MG to CBESS and grid; (**c**) buying power of MG from CBESS and grid; (**d**) centralized-based operation of CBESS.

By using the proposed method, the CBESS learns more and more to update its experience. After each episode, the CBESS updates its knowledge (Q-table) about the environment, as shown in Algorithm 2. Through the process, CBESS becomes more intelligent. The process is off-line training; it means the CBESS is trained with 24-hours forecasted data and the CBESS can optimally operate with its experience in real-time. After learning with a large number of episodes, CBESS can operate in an optimal way during a day based on its experience. Figure 7 shows the state value of CBESS (Q value) after learning with a different number of episodes. The red color represents the higher value of reward and the blue color represents a lower value of total reward for CBESS. The main idea of Q-learning is to try to receive the highest cumulative reward. With 10 episodes, CBESS randomly charges/discharges from the initial state (interval t = 1 and SoC = 50%), as shown in Figure 7a, when the episodes are increased to 1000 and 10,000, CBESS is more intelligent, as shown in Figure 7b,c. It tries to charge during off-peak price intervals (1–7) and discharges during peak price intervals (9–15 and 20–22). However, the CBESS also does not charge to 100% during off-peak price intervals. It is not an optimal way for operation of CBESS. Figure 7d shows the state values of CBESS with 50,000 episodes. CBESS charged/discharged almost similarly with the optimal operation that is obtained from the centralized-based operation. The detailed Q-learning-based CBESS actions are shown in Table 2. In order to clearly show the operation of CBESS, Figure 8 shows the operation of CBESS with a different number of episodes. It can be observed that the operation of CBESS converges to that of the centralized-based operation with an increase in the number of episodes. To determine the number of episodes, the CBESS is trained with a different number of episodes. When the results are not changed with a higher number of episodes, it means the model has achieved the optimal result. In this case study, the CBESS can reach the optimal operation with 100,000 episodes.



(**a**) (**b**)

(**c**) (**d**)

Figure 7. The state-action values (Q values) of community battery storage (CBES) with different number of episodes: (**a**) 10 episodes; (**b**) 1000 episodes; (**c**) 10,000 episodes; (**d**) 50,000 episodes.

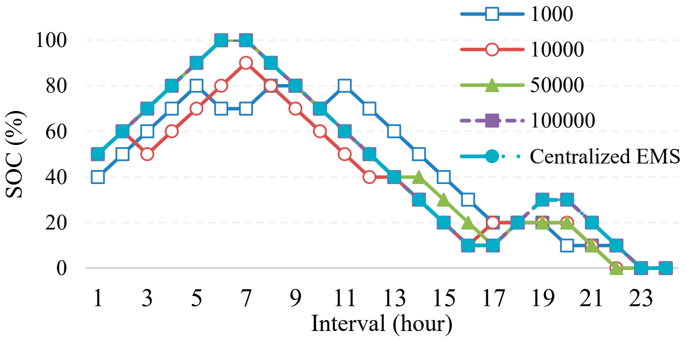


Figure 8. The operation of CBESS based on the Q-table with a different number of episodes.

Table 2. CBESS actions with 50,000 episodes.

Interval	**SoC**	**Action**	**Interval**	**SoC**	**Action**
	Initial = 50%				
1	50	Idle	13	50	Discharge
2	50	Charge	14	40	Idle
3	60	Charge	15	40	Discharge
4	70	Charge	16	30	Discharge
5	80	Charge	17	20	Discharge
6	90	Charge	18	10	Charge
7	100	Idle	19	20	Idle
8	100	Discharge	20	20	Idle
9	90	Discharge	21	20	Discharge
10	80	Discharge	22	10	Discharge
11	70	Discharge	23	0	Idle
12	60	Discharge	24	0	Idle

Finally, the total profit of CBESS is summarized in Table 3 with a different number of episodes. Total profit of CBESS is increased with the increase in a number of episodes. With 100,000 episodes, CBESS can get the highest profit and optimally operate like using a centralized EMS.

Table 3. Total profit of CBESS with the different number of episodes and centralized EMS.

Number of Episodes	1000	10,000	50,000	100,000	Centralized EMS
Total profit	25,640	26,850	27,420	27,990	27,990

3.3. Operation of the System in Islanded Mode

In islanded mode, the MG cannot trade with the utility grid. CBESS plays an important role in the operation of MGs for reducing the load shedding amount. CBESS is used to shift the surplus power to other intervals having shortage power. It means CBESS charges the surplus power and discharges for fulfilling the shortage power. Figure 9a shows the total surplus/shortage amount in the MG for each interval. The optimal operation of BESS is demonstrated in Figure 9b using the centralized EMS. The load shedding amount can be significantly decreased by interacting with the CBESS. The total load shedding amount is 597 kWh without CBESS and 306 kWh with CBESS, as shown in Figure 9c.

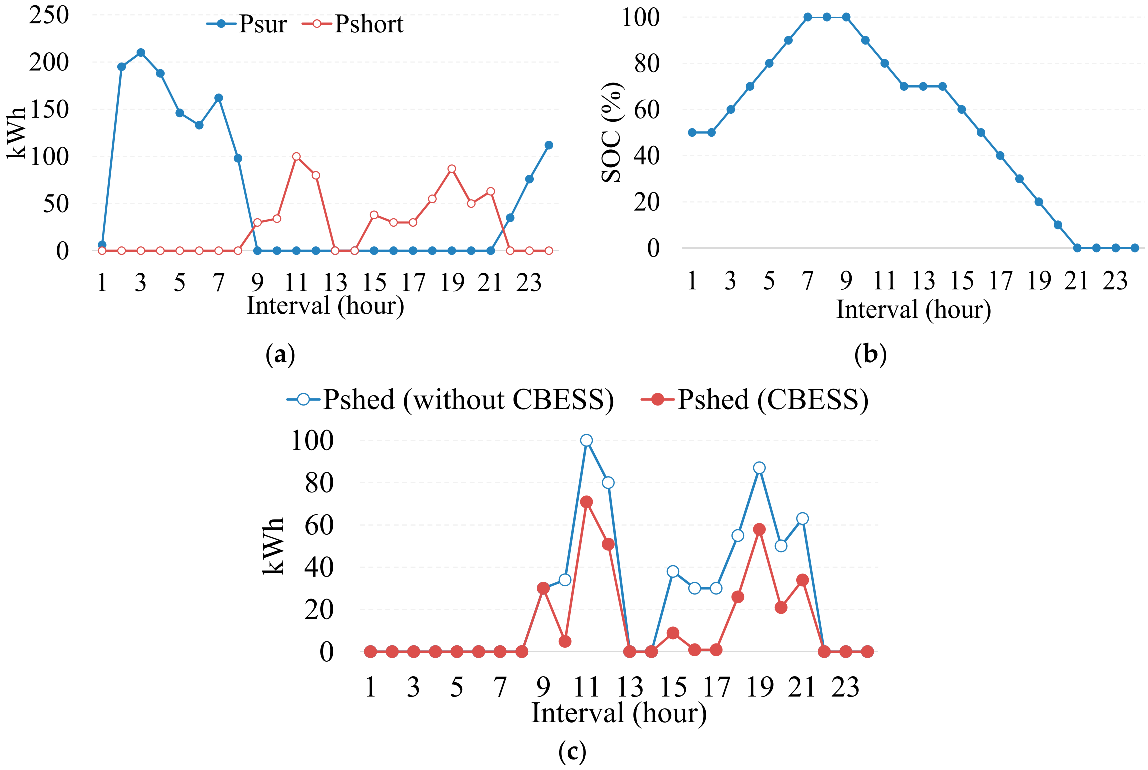


Figure 9. Operation of MG and CBESS in islanded mode: (**a**) the amount of surplus/shortage power in MG; (**b**) the optimal operation of CBESS by centralized EMS; (**c**) the load shedding amount in MG.

The Q-table of CBESS is trained by varying the number of episodes. The numerical results of Q-learning with a different number of episodes are also compared with the results of centralized-based operation, as shown in Figure 10. CBESS can optimize its operation similar to the centralized-based operation when the number of episodes is greater than 50,000. Table 4 shows the increase in load shedding amount compared with using the centralized EMS. The load shedding amount is reduced by increasing the number of episodes.

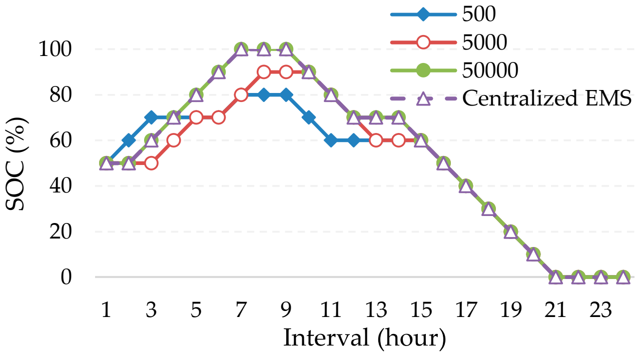


Figure 10. SoC of CBESS based on the Q-table with different episodes.

Table 4. Increasing load shedding amount with a different number of episodes and centralized EMS.

Number of Episodes	500	5000	50,000	Centralized EMS
Increasing in load shedding amount (%)	26.5	16.7	0	0

3.4. *Effects of Epsilon (ε) on the Operation of CBESS*

In this section, the effects of epsilon-greedy policy on the operation of CBESS are analyzed in detail. The value of epsilon is usually taken as a fixed value. It decides the probability of choosing a random action that is ε. To reduce the learning time, the value of epsilon is chosen as a lower value. However, it could get a trap in local optima. To overcome the problem, the value of epsilon is adjusted after some episodes for reducing the learning time and avoiding the trapping in local optima, as shown in Figure 11. When CBESS starts to learn, its knowledge about the environment is limited. Thus the value of epsilon is taken as a high value for increasing the exploring time. After each episode, CBESS has more knowledge about the environment, and the value of epsilon is decreased for using its prior knowledge.

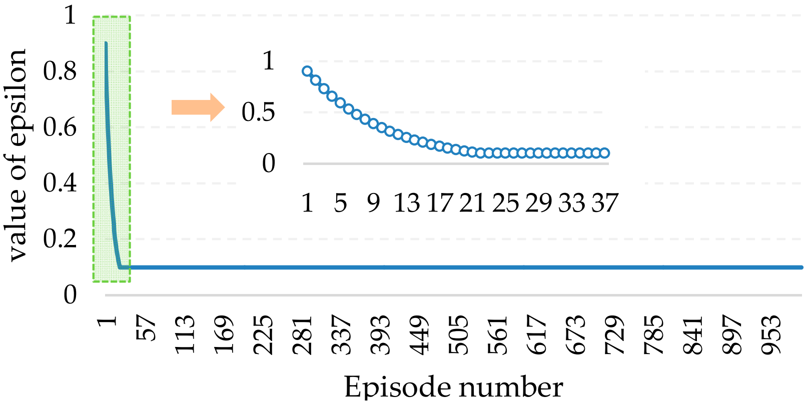
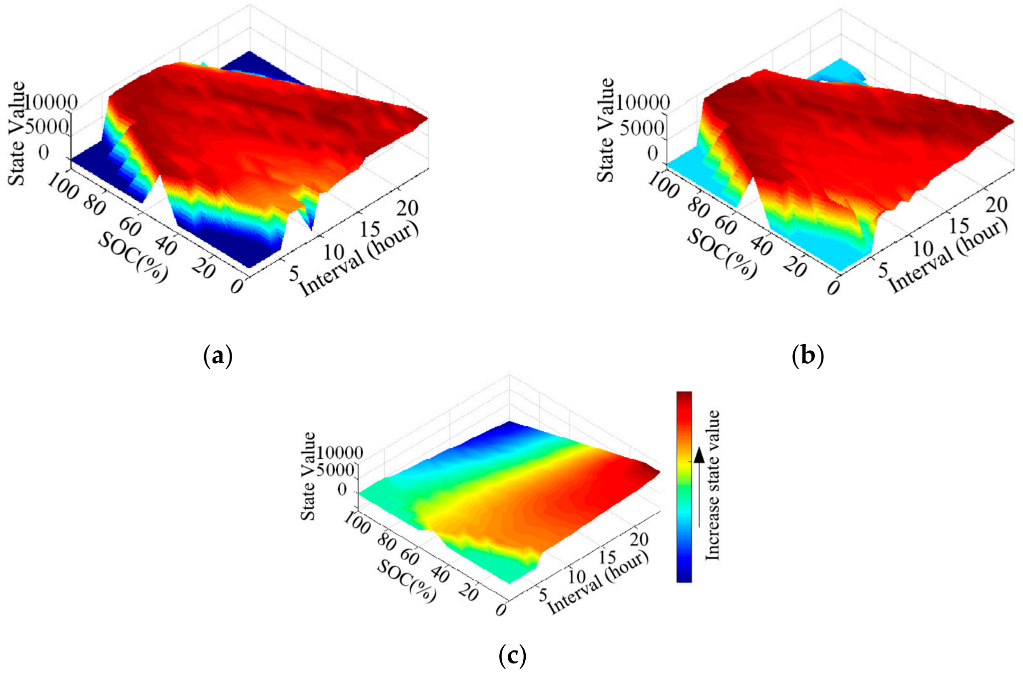


Figure 11. The value of epsilon during learning time.

Figure 12 shows the Q values of CBESS with different values of epsilon. It can be observed that the optimal operation of CBESS can be accurately determined when epsilon equals 0.9 or using the adjusted epsilon method. In case epsilon equals 0.1, CBESS got a local optimization, and CBESS always tried to discharge from the initial state (interval t = 1, SoC = 50%) to get its profit, but it is not the optimal way for operation of CBESS. Table 5 shows the learning time of CBESS with different values of epsilon. In case epsilon equals 0.9 or adjusted epsilon, the learning time is increased compared with the case of epsilon equals 0.1. However, the optimal operation can be accurately determined for both cases. Considering ε value 0.1 as the reference, the computation time increased by 13% and 8.9% for fixed epsilon (ε = 0.9) and adjusted epsilon, respectively. It can be concluded that the adjusted epsilon is the best way to reduce the leaning time with optimal results.



(**a**)

(**b**)

(**c**)

Figure 12. The state-action values (Q values) with different values of epsilon: (**a**) epsilon = 0.9; (**b**) adjustable epsilon; (**c**) epsilon = 0.1.

Table 5. Increasing learning time with a different value of epsilon.

Epsilon	0.9	0.1	Adjusted Epsilon
Computation time (s)	45.77	40.5	44.1
Increasing time (%)	13	0	8.9

4. Future Extension of the Proposed Strategy

In this study, an operation strategy for CBESS is proposed for maximizing profit in the grid-connected mode and reducing load shedding in islanded mode. The operation of CBESS is determined based on the information of the forecasted market price signals and surplus/shortage power of the MG. However, it is difficult to determine this information exactly. Thus the operation strategy of CBESS considering the uncertainties of market price signals and surplus/shortage power should be considered. These uncertainties result in a large state space, i.e., continuous state space.

As discussed in the previous section, in Q-learning, a Q-table is used to represent the knowledge of the agent about the environment. The Q-value for each state and action pair reflects the future reward associated with taking such an action in this state. However, Q-learning-based operation methods are only suitable for problems with small state space. They are not suitable for a continuous state space or for an environment with uncertainties. With any new state, the agent has to learn again to update the Q-table for optimizing the decisions. This could take a long time for the learning process in a real-time problem. Therefore, a model which maps the state information provided as input to Q-values of the possible set of actions should be developed, i.e., the Q-function approximator [28–30]. To solve this problem, Q-learning is combined with a deep neural network, which is called deep Q-learning method to enhance the performance of Q-learning for large scale problems. The operation strategy of CBESS using deep Q-learning will be discussed in a future extension of this study considering a continuous state space.

5. Conclusions

In this paper, a Q-learning-based energy management strategy has been developed for managing the operation of an MG integrated with a CBESS. An MG-EMS has been developed to manage the operation of the MG system. A Q-learning-based operation strategy for CBESS has been developed for optimizing its operation. By using the proposed strategy for CBESS, the efficiency and reliability of the entire system have significantly improved. Moreover, a comparison between the proposed strategy and a centralized-based method has been presented for showing the effectiveness of the proposed method. It can be observed that the CBESS can optimally work with the proposed strategy with a large number of episodes. The CBESS accurately determined the optimal operation like the centralized-based method in both grid-connected and islanded modes with different operation objectives. To reduce learning time, an adjusted epsilon has also been introduced for epsilon-greedy policy. By using the adjusted epsilon, the learning time has been reduced, and operation results have been improved.

Author Contributions: V.-H.B. conceived and designed the experiments; A.H. performed the experiments and analyzed the data; H.-M.K. revised and analyzed the results; V.-H.B. wrote the paper.

Funding: This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of South Korea (No. 20168530050030).

Acknowledgments: This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of South Korea (No. 20168530050030).

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

Nomenclature

Abbreviations	
BESS	Battery energy storage system
CDG	Controllable distributed generator
EMS	Energy management system
MG	Microgrid
RDG	Renewable distributed generator
SoC	State-of-charge
Identifiers and Sets	
I	Set of CDGs in MG
L	Set of residential loads in MG
T	Set of time intervals
Constants	
$C_i^{CDG}, C_i^{SU}, C_i^{SD}$	Operation, start-up, and shut-down costs of unit *i*
$P_i^{\min}, P_i^{\max}$	Operation bounds of unit *i*
PR_t^{Buy}, PR_t^{Sell}	Buying/selling price at time *t*
C_t^{pen}, C_t^{trade}	Penalty for power shortage and trading price in islanded mode at time *t*
RU_i, RD_i	Ramp up, ramp down rate of unit *i*
P_t^{PV}, P_t^{WT}	Output power of solar and wind turbine at time *t*
$P_{l,t}^{Load}$	Load amount of unit *l* at time *t*
P_B^{Cap}, P_{CB}^{Cap}	Capacity of BESS and CBESS
$SoC_{ini}^{B}, SoC_{ini}^{CB}$	Initial value of SoC of BESS and CBESS
$SoC_{\min}^{B}, SoC_{\max}^{B}$	Operation bounds of BESS
$SoC_{\min}^{CB}, SoC_{\max}^{CB}$	Operation bounds of CBESS
L^{B+}, L^{B-}	Losses for charging/discharging of BESS
L^{CB+}, L^{CB-}	Losses for charging/discharging of CBESS
Variables	
$u_{i,t}$	On/off mode of unit *i* at time *t*
$y_{i,t}, z_{i,t}$	Start-up and shut-down statuses of unit *i* at time *t*
$P_{i,t}^{CDG}$	Output power of unit i at time *t*

P_t^{Buy}, P_t^{Sell}	Buying/selling power at time t from/to the external systems
$P_t^{Sell\_Grid}$	Selling power at time t from MG to the utility grid
$P_t^{Sell\_CBESS}$	Selling power at time t from MG to CBESS
$P_t^{Buy\_Grid}$	Buying power of MG at time t from the utility grid
$P_t^{Buy\_CBESS}$	Buying power of MG at time t from CBESS
$P_{Grid,t}^{CB+}, P_{MG,t}^{CB+}$	Charging power of CBESS at time t from the utility grid and MG
$P_{Grid,t}^{CB-}, P_{MG,t}^{CB-}$	Discharging power of CBESS at time t to the utility grid and MG
P_t^{B+}, P_t^{B-}	Charging/discharging power of BESS at time t.
SoC_t^B	State of charge of BESS at time t
P_t^{CB+}, P_t^{CB-}	Charging/discharging power of CBESS at time t.
SoC_t^{CB}	State of charge of CBESS at time t
$Q(s,a)$	Q value of state s, doing action a
P_t^{Sur}, P_t^{Short}	Surplus/shortage power of MG at time t

References

1. Chen, C.; Duan, S.; Cai, T.; Liu, B.; Hu, G. Smart energy management system for optimal microgrid economic operation. *IET Renew. Power Gener.* **2011**, *5*, 258–267. [CrossRef]
2. Jiang, Q.; Xue, M.; Geng, G. Energy management of microgrid in grid-connected and stand-alone modes. *IEEE Trans. Power Syst.* **2013**, *28*, 3380–3389. [CrossRef]
3. Katiraei, F.; Iravani, R.; Hatziargyriou, N.; Dimeas, A. Microgrid management. *IEEE Power Energy Mag.* **2008**, *6*, 54–65. [CrossRef]
4. Su, W.; Wang, J. Energy management systems in microgrid operations. *Electr. J.* **2012**, *25*, 45–60. [CrossRef]
5. Olivares, D.E.; Cañizares, C.A.; Kazerani, M. A centralized energy management system for isolated microgrids. *IEEE Trans. Smart Grid* **2014**, *5*, 1864–1875. [CrossRef]
6. Vaccaro, A.; Loia, V.; Formato, G.; Wall, P.; Terzija, V. A selforganizing architecture for decentralized smart microgrids synchronization, control, and monitoring. *IEEE Trans. Ind. Inform.* **2015**, *11*, 289–298. [CrossRef]
7. Hussain, A.; Bui, V.H.; Kim, H.M. A resilient and privacy-preserving energy management strategy for networked microgrids. *IEEE Trans. Smart Grid* **2018**, *9*, 2127–2139. [CrossRef]
8. Kim, H.-M.; Kinoshita, T.; Shin, M.-C. A multiagent system for autonomous operation of islanded microgrids based on a power market environment. *Energies* **2010**, *3*, 1972–1990. [CrossRef]
9. Wang, Z.; Chen, B.; Wang, J.; Kim, J. Decentralized energy management system for networked microgrids in grid-connected and islanded modes. *IEEE Trans. Smart Grid* **2016**, *7*, 1097–1105. [CrossRef]
10. Harmouch, F.Z.; Krami, N.; Hmina, N. A multiagent based decentralized energy management system for power exchange minimization in microgrid cluster. *Sustain. Cities Soc.* **2018**, *40*, 416–427. [CrossRef]
11. Bui, V.H.; Hussain, A.; Kim, H.M. A multiagent-based hierarchical energy management strategy for multi-microgrids considering adjustable power and demand response. *IEEE Trans. Smart Grid* **2018**, *9*, 1323–1333. [CrossRef]
12. Kara, E.C.; Berges, M.; Krogh, B.; Kar, S. Using smart devices for system-level management and control in the smart grid: A reinforcement learning framework. In Proceedings of the IEEE Smart Grid Communications (SmartGridComm), Tainan, Taiwan, 5–8 November 2012; pp. 85–90.
13. Dimeas, A.L.; Hatziargyriou, N.D. Multi-agent reinforcement learning for microgrids. In Proceedings of the IEEE Power and Energy Society General Meeting, Minneapolis, MN, USA, 25–29 July 2010; pp. 1–8.
14. Dimeas, A.L.; Hatziargyriou, N.D. Agent based control for microgrids. In Proceedings of the IEEE Power Engineering Society General Meeting, Tampa, FL, USA, 24–28 June 2007; pp. 1–5.
15. Wei, C.; Zhang, Z.; Qiao, W.; Qu, L. An adaptive network-based reinforcement learning method for MPPT control of PMSG wind energy conversion systems. *IEEE Trans. Power Electron.* **2016**, *31*, 7837–7848. [CrossRef]
16. Venayagamoorthy, G.K.; Sharma, R.K.; Gautam, P.K.; Ahmadi, A. Dynamic energy management system for a smart microgrid. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 1643–1656. [CrossRef] [PubMed]
17. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; The MIT Press: London, UK, 2018; pp. 1–398.
18. Panait, L.; Luke, S. Cooperative multi-agent learning: The state of the art. *Auton. Agents Multi-Agent Syst.* **2005**, *11*, 387–434. [CrossRef]

19. Mbuwir, B.V.; Ruelens, F.; Spiessens, F.; Deconinck, G. Battery energy management in a microgrid using batch reinforcement learning. *Energies* **2017**, *10*, 1846. [CrossRef]
20. Kim, S.; Lim, H. Reinforcement learning based energy management algorithm for smart energy buildings. *Energies* **2018**, *11*, 2010. [CrossRef]
21. Leo, R.; Milton, R.S.; Kaviya, A. Multi agent reinforcement learning based distributed optimization of solar microgrid. In Proceedings of the IEEE Computational Intelligence and Computing Research, Coimbatore, India, 18–20 December 2014; pp. 1–7.
22. Kuznetsova, E.; Li, Y.F.; Ruiz, C.; Zio, E.; Ault, G.; Bell, K. Reinforcement learning for microgrid energy management. *Energy* **2013**, *59*, 133–146. [CrossRef]
23. Li, F.D.; Wu, M.; He, Y.; Chen, X. Optimal control in microgrid using multi-agent reinforcement learning. *ISA Trans.* **2012**, *51*, 743–751. [CrossRef]
24. Wiering, M.; Ma, M.O. *Reinforcement Learning State-of-the Art*; Springer: Berlin/Heidelberg, Germany, 2012.
25. Mohan, Y.; Ponnambalam, S.G.; Inayat-Hussain, J.I. A comparative study of policies in Q-learning for foraging tasks. In Proceedings of the IEEE Nature & Biologically Inspired Computing, Coimbatore, India, 9–11 December 2009; pp. 134–139.
26. *IBM ILOG CPLEX V12.6 User's Manual for CPLEX 2015, CPLEX Division*; ILOG: Incline Village, NV, USA, 2015.
27. Lee, H.; Byeon, G.S.; Jeon, J.H.; Hussain, A.; Kim, H.M.; Rousis, A.O.; Strbac, G. An energy management system with optimum reserve power procurement function for microgrid resilience improvement. *IEEE Access* **2019**, *7*, 42577–42585. [CrossRef]
28. Radac, M.B.; Precup, R.E.; Roman, R.C. Data-driven model reference control of MIMO vertical tank systems with model-free VRFT and Q-Learning. *ISA Trans.* **2018**, *73*, 227–238. [CrossRef]
29. Kofinas, P.; Dounis, A.I.; Vouros, G.A. Fuzzy Q-Learning for multi-agent decentralized energy management in microgrids. *Appl. Energy* **2018**, *219*, 53–67. [CrossRef]
30. Vázquez-Canteli, J.R.; Nagy, Z. Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Appl. Energy* **2019**, *235*, 1072–1089. [CrossRef]