IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 13, 2020, accepted February 2, 2020, date of publication February 17, 2020, date of current version March 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974286

Demand Response Strategy Based on Reinforcement Learning and Fuzzy Reasoning for Home Energy Management

FAYIZ ALFAVERH, M. DENAÏ, AND YICHUANG SUN, (Senior Member, IEEE)
School of Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, U.K.
Corresponding author: M. Denaï (m.denai@herts.ac.uk)

ABSTRACT As energy demand continues to increase, demand response (DR) programs in the electricity distribution grid are gaining momentum and their adoption is set to grow gradually over the years ahead. Demand response schemes seek to incentivise consumers to use green energy and reduce their electricity usage during peak periods which helps support grid balancing of supply-demand and generate revenue by selling surplus of energy back to the grid. This paper proposes an effective energy management system for residential demand response using Reinforcement Learning (RL) and Fuzzy Reasoning (FR). RL is considered as a model-free control strategy which learns from the interaction with its environment by performing actions and evaluating the results. The proposed algorithm considers human preference by directly integrating user feedback into its control logic using fuzzy reasoning as reward functions. Q-learning, a RL strategy based on a reward mechanism, is used to make optimal decisions to schedule the operation of smart home appliances by shifting controllable appliances from peak periods, when electricity prices are high, to off-peak hours, when electricity prices are lower without affecting the customer's preferences. The proposed approach works with a single agent to control 14 household appliances and uses a reduced number of state-action pairs and fuzzy logic for rewards functions to evaluate an action taken for a certain state. The simulation results show that the proposed appliances scheduling approach can smooth the power consumption profile and minimise the electricity cost while considering user's preferences, user's feedbacks on each action taken and his/her preference settings. A user-interface is developed in MATLAB/Simulink for the Home Energy Management System (HEMS) to demonstrate the proposed DR scheme. The simulation tool includes features such as smart appliances, electricity pricing signals, smart meters, solar photovoltaic generation, battery energy storage, electric vehicle and grid supply.

INDEX TERMS Demand response, home energy management system, smart home, smart appliances, reinforcement learning, Q-learning, fuzzy reasoning.

I. INTRODUCTION

Greenhouse gas emissions are posing a serious concern across the world due to their negative impacts on the environment and climate change. On the other hand, the global economy is in the midst of unprecedented demand for energy seeking new investments for the reinforcement and expansion of power grid infrastructures and the large adoption of renewable energy resources. As a result, the electric power sector around the world is experiencing an ongoing global restructuration to establish the ground rules and legislations for the generation and trading of electricity from this energy mix. This has created deregulated wholesale electricity markets, mostly in developed countries, and the emergence of new business opportunities for independent producers and energy service providers which are changing the way energy is bought and sold. A reliable operation of the electricity grid, under these conditions, requires that supply and demand must be perfectly balanced [1], [2].

DR programs are being introduced by some electricity grid operators as resource options for curtailing and reducing the demand of electricity during certain time periods for

The associate editor coordinating the review of this manuscript and approving it for publication was Jihwan P. Choi.

balancing supply and demand. DR is considered as a class of demand-side management programs, where utilities offer incentives to end-users to reduce their power consumption during peak periods [3]. DR is, indeed, a promising opportunity for consumers to control their energy usage in response to electricity tariffs or other incentives from their energy suppliers [4], [5].

Generally, DR schemes are classified into two categories namely incentives-based programs and price-based programs. In incentives-based programs, participants receive fix or time-varying payments for their consent to reduce power consumption during peak demand or system contingencies. There are two categories: classical programs and market-based programs. Participating in classical programs offers participation payments as bill credits or discount rates. In market-based programs, customers receive money rewards depending on their performance after they consent to reduce their power consumption during peak periods [6].

Incentive-based programs include: Direct Load Control (DLC), Interruptible/Curtailable (I/C) and Emergency DR programs. DLC programs are considered as classical incentive-based programs. They enable utility companies to remotely turn off consumers' electrical loads. Participants in this program receive payments in return for reducing their energy usage below a pre-defined threshold. In I/C DR programs, participants are also offered economic incentives. The power utility can curtail a specific part or the total users' consumption to a certain level during emergency situations. Consumers who do not reduce their energy consumption receive penalties as per the pre-defined terms and conditions of the program. Emergency DR program are a combination of both DLC and I/C programs and are considered as market-based programs.

Price-based programs, on the other hand, can be considered as indirect means for controlling customers' loads. Using these programs, time-varying prices are offered to customers based on electricity cost at different time periods. Customers willing to reduce their energy usage during peak hours, when the electricity prices are high, can participate in these programs. They are expected to adjust their demand in response to electricity price signals [7]. Price-based programs are of three types: Time-of Use (TOU) pricing, Real-Time-Pricing (RTP) and Inclining Block Rate (IBR).

In TOU tariff plan, electricity pricing varies depending on the time of the day, day of the week and season. It contains three time periods namely; off-peak, mid-peak and on-peak period. TOU pricing is easy to follow and give participants the opportunity to take control of their energy usage by shifting their electricity consumption to lower-prices hours. While TOU pricing reduces the electricity demand during peak hours, there is a risk that this may create a similar or larger peak demand during off-peak periods [8]. Under RTP, electricity prices change over short time periods typically hourly or less and are announced in advance by energy suppliers. The IBR program has a two-level rate structure with lower and higher electricity price. It aims to incentivise users to avoid high prices by distributing their consumption across different periods of the day.

Home Energy Management System (HEMS) provides the interface for consumers to monitor and control their various household electrical devices in real-time. HEMS can be considered as the enabling technology for realizing the potential of DR strategies and enable consumers to improve the energy usage and minimise electricity bills by shifting and curtailing their loads in response to electricity tariffs during peak periods without compromising their lifestyle and preferences [3], [5], [9], [10].

User's comfort has mainly been considered in HEMS. In [11], the authors proposed a scheduling model for HEMS considering energy payment and user's preferences level as a comprehensive objective in the optimization process. The HEMS is proposed in [12] with the objective to reduce the electricity cost and avoid compromising consumers' lifestyle and preferences. The authors in [13] focused on HEMS algorithm considering customer preferences setting, priority of appliances and comfortable lifestyle.

Although HEMS technology is still in its early stages, in the past few years, the market for HEMS has been on the rise and is quickly expanding. Many researchers have worked on developing HEMS using rule-based control strategies. In [14], the author proposed a Hybrid Genetic Particle Swarm Optimisation (HGPO) to schedule the appliances of a house with local generation from Renewable Energy Sources (RES). However, this algorithm attempts to minimise electricity bills without considering consumer's preferences. Optimisation techniques based on Integer Linear Programming (ILP) and Dynamic Programming (DP) have been used to manage energy usage and reduce the electricity cost in smart homes. In [15], the household appliances are divided into two types; appliances with a flexible starting time and a fixed power, and other appliances with a flexible power and a predefined working time. This approach aimed to achieve a desired trade-off between electricity bills reduction and discomfort where the users can modify the starting time of the first type of appliances or reduce the energy consumption of the second appliances to reduce the bills. However, this algorithm does not consider consumer's comfort. The authors in [16] focused on load scheduling problems and power trading using DP algorithm. This enables users to sell their surplus of generated power to the power grid or other local users. However, due to its computational complexity, the model is difficult to implement in real-time.

Recently, much attention has been devoted to the development of controllers based on computational intelligence and machine learning techniques for HEMS [17], [18]. According to RTP program, end-users receive energy prices from power utility an hour-ahead in order to make a decision to shift or reduce their energy consumption. Therefore, in [18], Artificial Neural Networks (ANN) have been used to design energy price forecasting models and overcome the uncertainty in future prices. The ANN approach is used due to its ease of implementation, good performance and less time-consuming.

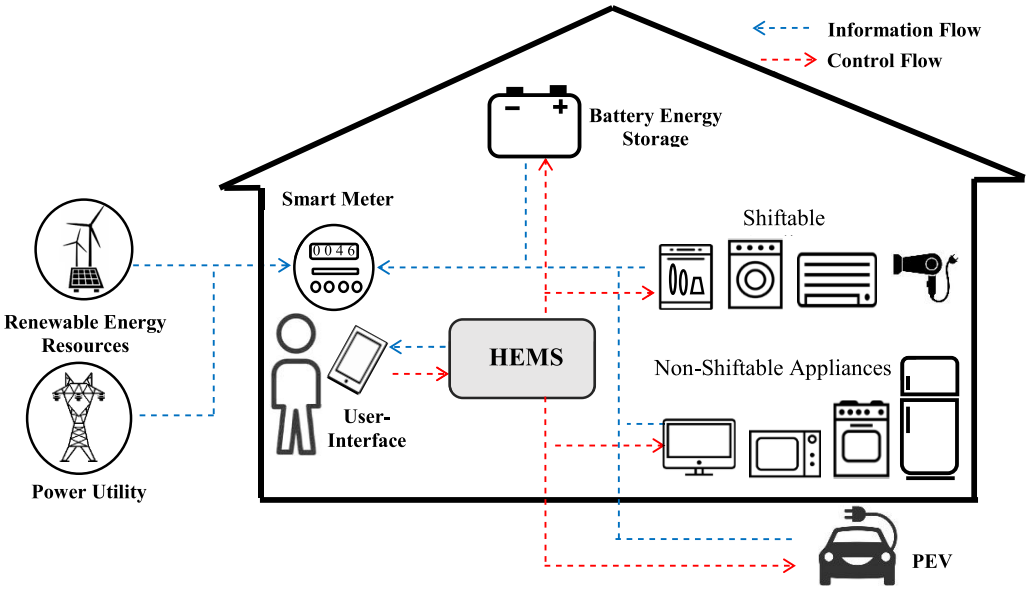

FIGURE 1. Smart HEMS architecture including Renewable Energy Resources, Energy Storage (battery), Power Utility, User-interface, Smart HEMS Center, and Household Appliances (Shiftable and Non-shiftable).

Recently, Reinforcement Learning (RL) has emerged as a potential machine learning algorithm for energy management, decision and control. RL models have excellent decision-making ability due to their potential to solve problems without a priori knowledge of the environment. Multi-agent reinforcement learning has been proposed for the optimal scheduling of household appliances to optimise the energy utilisation [17], [19]. However, multi-agents RL requires setting several agents, where each household appliances represents an environment that has its own agent with different actions and rewards. Therefore, the learning process becomes more complex [20]. Other studies have focused on using Q-learning and SARSA (State-Action-Reward-State-Action) algorithms in HEMS to schedule controllable appliances and shift the operation time of shiftable devices [21], [22]. However, these algorithms require many state-action pairs and consequently the convergence speed of the Q-values is reduced. In this research, a new and flexible HEMS is proposed, to smooth the power consumption profile without compromising user comfort and preferences. The proposed approach works with a single agent and uses a reduced number of state-action pairs and fuzzy logic for rewards functions.

This paper is organised as follows: In Section II, the HEMS architecture and functionalities are briefly described. In Section III, the concepts of RL and Q-learning are overviewed. HEMS and RL models are presented in Section IV. Section V presents the results and discussion. Finally, the conclusions of the paper are summarised in Section VI.

II. DESCRIPTION OF THE HEMS ARCHITECHTURE AND FUNCTIONALITIES

Smart HEMS is an essential home system to achieve an effective demand-side response (DSR) and DR in the context of smart grids. It is used to monitor, control and optimise the amount of energy consumed or to be consumed in real time, based on the customer's preferences via a Human-Machine Interface (HMI). Consequently, this helps users to actively participate in DR programs to reduce electricity cost and achieve efficient energy utilisation by shifting electricity consumption during peak demand in response to changes in the electricity price. To achieve electricity saving and DR objectives, HEMS should be more flexible and able to manage different types of household resources such as Renewable Energy Sources (RERs) and Home Energy Storage System (HESS). Power consumption and electricity pricing should be offered to users in real-time to enables them choose their preferences to schedule the operation time of various appliances via the HMI which in turn improves their energy usage efficiency.

A. SMART HEMS ARCHITECTURE

HEMS will play an integral role in future smart electricity networks. They provide end-users with the ability to participate in demand response which aims to optimise energy utilisation and minimise electricity bills.

Figure 1 illustrates a typical smart HEMS architecture. The system includes a user interface, smart meters, home communication networks and smart household appliances. Smart meters are advanced energy electricity meters which offer, in real-time, a range of services to households, such as information about electricity usage, local generation from RER and costs, via a two-way communication infrastructure Since each household appliance has a specific electrical characteristic and energy consumption profile, several studies have focused on the disaggregation of the whole home energy profiles into appliance-by-appliance energy usage profile. Energy disaggregation, also known as Non-Intrusive Load Monitoring (NILM), takes the total energy consumption and

attempts to match the disaggregated signals to individual appliances. In [3], a NILM based on deep learning techniques is developed and tested. The algorithm can identify household electrical appliances and their energy consumed using smart meter measurement. However, NILM techniques tend to reveal consumer's habits and life style and presents privacy concerns. Therefore, several researchers have worked on the privacy preserving techniques of smart meters. In [23], a real-time different privacy load monitoring (DPLM) algorithm is proposed using Laplacian Noise. A privacy-preserving and efficient data aggregation scheme is proposed in [24]. Authors divide users into different groups where each group has a private database to store his/her data. Pseudonyms to hide costumers' identities are used to preserve the privacy of each group.

In the last decade, different communication and network technologies for HEMS have been designed to connect smart devices with each other and exchange information to allow users to remotely manage and control their devices. Recently, many protocols have been used in Home Area Networks (HANs), such as Bluetooth, ZigBee, BACnet and INSTEON. Small-scale networks (12 to 100 meters) such as Local Area Network (LAN), Body Area Network (BAN) and Personal Area Network (PAN) are integrated to HEMS to provide users with movement flexibility and do not need high expertise to manage the network operations. In [25], [26], ZigBee protocol with PAN is used for the proposed HEMS. ZigBee is considered as a low power, low cost wireless communication technology for HEMS.

Household appliances are usually classified into shiftable and non-shiftable. Where shiftable refer to the class of appliances that can operate at any time within user's defined time periods (such as washing machine, dishwasher and clothes dryer). Non-shiftable refer to appliances that require permanent electric power supply to complete their tasks (such as refrigerator, water heater and lighting). An additional class of appliances includes battery-assisted devices. In [27], major home appliances, such as dishwasher, clothes washer and dryer, refrigerator, air-conditioning and oven are described.

B. SMART HEMS FUNCTIONALITIES

The primary aim of a smart HEMS is to provide efficient management and control systems to achieve the DR objectives. Therefore, it should be flexible enough to manage several power consumption patterns, dynamic electricity prices and different types of household appliances. HEMS enables consumers easy access to their energy usage data in real-time to make them more aware about their electricity saving. It also provides services for the operational modes and energy status of each household appliance via HMI.

The control functionality provides customers the ability to access their household appliances and can be classified into two types namely, direct control and remote control. Whereas remote control enables consumers to monitor and control their appliances on-line via a personal computer or smart phone from outside the home.

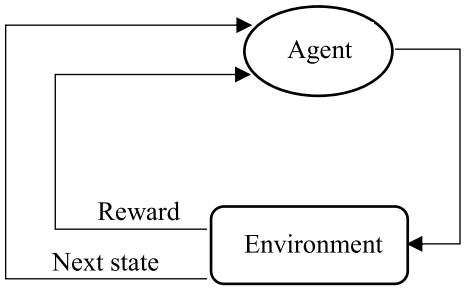

FIGURE 2. Reinforcement learning process.

The key function of HEMS is energy management services in order to optimise the power consumption in the smart home. This functionality includes renewable energy generation management, energy storage management, home appliance management.

HEMS also collect and store data on power consumption of appliances, generation from renewable energy resources, and energy storage state of charge. It also receives real-time prices from power utility and performs demand response analysis.

III. REINFORCEMENT LEARNING AND Q-VALUE

Household Energy Management (HEM) is an optimisation problem, which aims to minimise the total power consumption of electrical appliances and reduce the electricity bills in a smart home. A typical HEMS can neither be adapted to a variety of appliances with varying scheduling complexity nor it is appropriate for real-time application. Reinforcement Learning (RL) algorithms have been recently proposed as potential candidates to address these issues due to their adaptability and ability to learn customer's preferences, and optimise the management of energy systems which are often subject to various inputs such as dynamic electricity prices, forecast data and energy consumption patterns [28], [29]. RL is considered as a machine-learning type of algorithm for decision-making in a stochastic environment [10] . It does not require a mathematical model and is suitable for complex and real-time applications. RL algorithm has six parameters namely, agent, environment, state space S, action space A, rewards R, and action-value $Q(s, a)$. Generally, the RL-agent interacts with an environment as illustrated by Figure 2.

Firstly, at each time step $t = \{0, 1, 2, \ldots\}$, the agent executes an action according to a certain policy π at a current state $s_t \in S(t)$. The environment then computes the new state $s_{t+1} \in S(t)$ and a numerical reward $r(s_t, a_t)$ and feed it back to the agent in order to evaluate the action taken as shown in Figure 2. Based on the reward received, the agent is able to optimise its policy π and hence maximise the total rewards it will receive in the future.

The action-value function which indicates how good is the action taken in each state is denoted by $Q_\pi(s, a)$. According to a certain policy π, $Q_\pi(s, a)$ expresses the value of action taken a_t and is selected from a valid set of actions space A in the current state s_t:

$$Q_\pi(s, a) = E_\pi \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a \right] \quad (1)$$

E_π denotes the expectation of total rewards defined by policy π. γ is called the discount rate and indicates the relationship between the future and current rewards. It takes a fraction between [0, 1]. When $\gamma = 0$, the agent considers only the current reward, while $\gamma = 1$ means that the agent will strive for the future rewards. For each state, there is at least one optimal action which receives the highest reward. Therefore, the policy works to select the action with the highest Q-value as follows:

$$\pi\left(a \mid s\right) = argmaxQ\left(s, a\right) \tag{2}$$

Q-Learning algorithms are RL techniques that are adopted to acquire the optimal policy π. The main procedure of Q-Learning is to assign a Q-value $Q(s_t, a_t)$ to each state-action pair at time step t, and then update this value at each iteration in order to optimise the agent's performance. The optimal $Q^*_\pi\left(s_t, a_t\right)$ expresses the maximum discounted achieved with the future reward $r\left(s_t, a_t\right)$ for action a_t taken at state s_t, which is expressed as follows:

$$Q^*_\pi\left(s_t, a_t\right) = r\left(s_t, a_t\right) + \gamma \cdot \max Q\left(s_{t+1}, a_{t+1}\right) \tag{3}$$

Once the action a_t is taken based on a certain policy π, the defined reward $r\left(s_t, a_t\right)$ (or calculated using reward function) will be received, and then the agent assume a new state s_{t+1}. Simultaneously, the action-value $Q\left(s_t,a_t\right)$ is updated using the following equation:

$$Q\left(s_t, a_t\right) \leftarrow (1-\alpha) Q\left(s_t, a_t\right) + \alpha \begin{bmatrix} r\left(s_t, a_t\right) + \\ \gamma \cdot \max Q\left(s_{t+1}, a_{t+1}\right) \end{bmatrix} \tag{4}$$

where α denotes the learning rate which determines how much the new reward affects the old value of the $Q\left(s_t, a_t\right)$. For example, $\alpha = 0$ means that the new information acquired is not used in the leaning process and hence the reward received does not affect the Q-value. When $\alpha = 1$, only the latest information is considered.

IV. HOME ENERGY MANAGEMENT AND Q-LEARNING MODELLING

In this section, the HEM structure is presented, where RL is modelled using Q-learning algorithm that contains a state space, action space, reward definition.

A. HOME ENERGY MANAGEMENT MODEL

Figure 3 shows the daily power demand profile of a typical household. Two peak demand periods occur during morning and evening times when energy prices are higher. Whereas off-peak demand periods correspond to periods of the day where electricity prices are lower since customer's activities such as washing, cleaning, cooking, and watching TV are reduced [30]. Therefore, the aim of this study is to shift the operating time of specific appliances from peak demand hours to off-peak periods without compromising the costumer's preferences. In this study, household appliances are divided into shiftable and non-shiftable appliances.

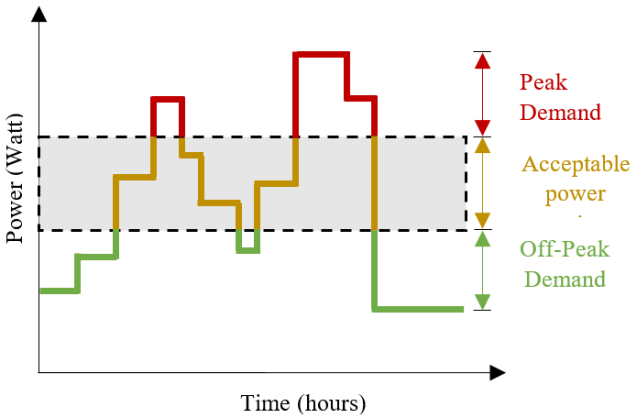

FIGURE 3. Daily household power demand.

TABLE 1. Rated power for non-shiftable appliances.

Index Number	Appliance	Rated Power (W)
1	Iron	1000
2	Oven	2000
3	Laptop	20
4	Microwave	600
5	Television	200
6	Lighting	100
7	Refrigerator	200
8	Water heater	2000

1) NON-SHIFTABLE APPLIANCES

Once started, these appliances must be continuously powered to complete their tasks and they cannot be shifted to another time regardless of the electricity price.

Table 1 shows the rated power consumption of non-shiftable household appliances. The total power consumption of these appliances at each time step is:

$$E_t^{non} = \sum_{n=1}^{N} e_t^{n,non} \cdot I_t^n \tag{5}$$

E_t^{non} represents the total power demand of all non-shiftable appliances for each hour, $e_t^{n,non}$ is the rated power of a specific non-shiftable appliance, I_t^n denotes the status of the appliance and takes values 0 (off) or 1 (on) respectively, $t \in \{1, 2, 3 \ldots 24\}$ represents the hour of the day, $n \in \{1, 2, \ldots N$ is the appliance number and N is the total number of the non-shifatble appliances.

2) SHIFATBLE APPLIANCES

Shiftable appliances include the washing machine, dish washer, electric vehicle and others, and their operation time can be re-scheduled based on appliance priority and preference setting. The power demand of these appliances is

TABLE 2. Rated power for non-shiftable appliances.

Index Number	Appliance	Rated Power (W)	Priority
1	Washing machine	800	1
2	Dish washer	1100	2
3	Clothes dryer	400	3
4	Hair dryer	450	4
5	Hair straightener	20	5
6	PEV	1200	6

defined as follows:

$$E_t^{sht} = \sum_{n=1}^{N} e_t^{n,s\,ht} \cdot I_t^n \tag{6}$$

where E_t^{sht} is the total power required from all shiftable appliances for each hour, $e_t^{n,sht}$ represents the rated power of each shiftable appliance at that hour. The rated power for shiftable appliances is illustrated in Table 2.

Therefore, at each time step (considered as an hour in this work), the total power demand of both shiftable and non-shiftable appliances during a certain hour is:

$$E_t^{total} = E_t^{non} + E_t^{sht} \tag{7}$$

3) DEMAND RESPONSE PROGRAM

Due to the changes in the electricity price during a day, DR program aims to inform customers about the prices on hour-ahead basis. Smart meters receive the RTP signal from the utility and record the current power demand data of all household appliances during their operating times, and then send them to the HEM system.

B. Q-LEARNING MODEL

RL is adopted to make an optimal decision in a stochastic environment (dynamic electricity prices and different energy consumption patterns) using an intelligent agent. Practically, the agent can control a dynamic system by executing sequential actions. Where the dynamic system could be characterised by a state-space and a numerical reward that evaluates the new state when a given action is taken. In this paper, the Q-learning model components are defined as follows:

1) STATE SPACE

The state-space here is represented by the power demand and the electricity price signal. To reduce the computation time and make the model much simpler, the power demand is divided into three levels namely; low, average and high-power demand. Whereas the price signal is categorised into cheap and expensive price as follows:

TABLE 3. Indexing of all possible states.

Power Demand	Price	State Index
E_{high}^{total}	$P_t^{expensive}$	6
E_{high}^{total}	P_t^{cheap}	5
$E_{average}^{total}$	$P_t^{expensive}$	4
$E_{average}^{total}$	P_t^{cheap}	3
E_{low}^{total}	$P_t^{expensive}$	2
E_{low}^{total}	P_t^{cheap}	1

$$E_{t,index}^{total} = \begin{cases} E_{low}^{total}, & \text{if } E_t^{total} \leq 3.8kW \\ E_{average}^{total}, & \text{if } 4.65W < E_t^{total} < 3.8kW \\ E_{high}^{total}, & \text{if } E_t^{total} \geq 4.65kW \end{cases} \tag{8}$$

$$P_t^{index} = \begin{cases} P_t^{cheap}, & \text{if } P_t \leq 0.1£/kWh \\ P_t^{expensive}, & \text{if } P_t > 0.1£/kWh \end{cases} \tag{9}$$

For each time step (an hour), the state is defined to contain both power demand and electricity price indexes:

$$s_t = \left[E_{t,index}^{total}, P_t^{index}\right] \tag{10}$$

Table 3 summarises all available states that can be created from power demand and real-time electricity price. It also shows the index of each state.

2) ACTION SPACE

The aim is to shift the operating time of the specific appliance that has the lowest priority during peak demand when required, and then turn on the appliance that has the highest priority during off-peak hours.

Based on the relationship of the real-time price, the total power demand of all household appliances, taking into account load priority and customer preferences, the agent (HEMS) chooses one action from the action space A that given by: -

$$A = [do\ nothing,\ shifting,\ valley\ filling] \tag{11}$$

where *shifting* action shifts the lowest priority device. This mode occurs always during peak demand when the price and the power consumed are high. *Valley-filling* action seeks to turn on the shifted appliance with the highest priority, usually during off-peak demand hours. When *do-nothing* is set, the system works in normal conditions and there is no need to shift any appliance.

3) REWARDS FUNCTION IMPLEMENTATION USING FUZZY LOGIC

Let $r(s_t, a_t)$ denote the numerical reward that the agent receives after executing a random action and observing a

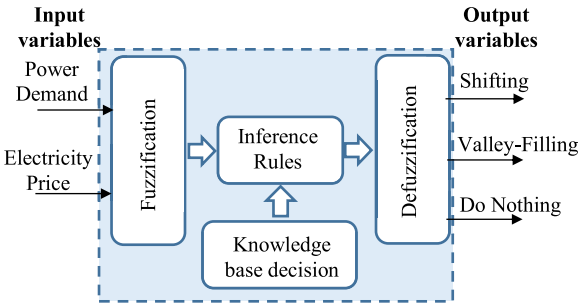

FIGURE 4. FIS system of the reward function.

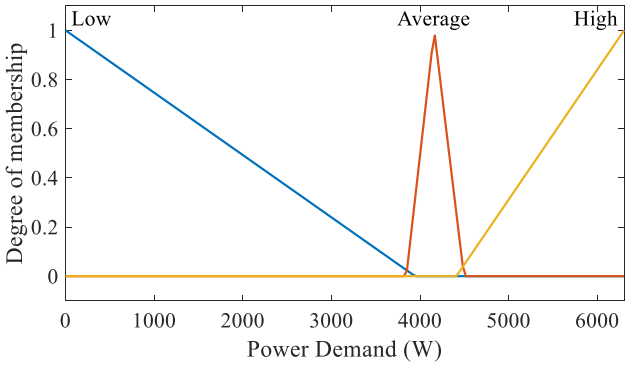

FIGURE 5. Fuzzy sets and MFs of power demand input.

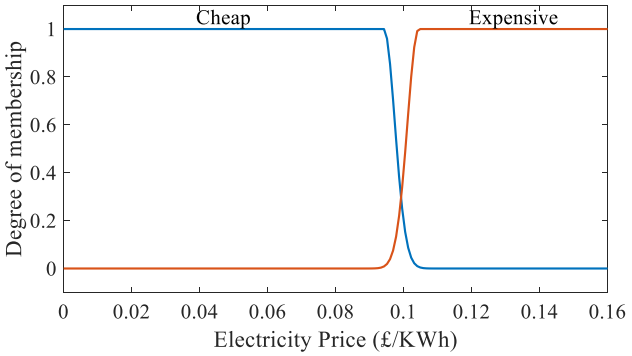

FIGURE 6. Fuzzy sets and MFs of electricity price input.

TABLE 4. Fuzzy rules of fis system.

Power Demand	Electricity Price	*shifting*	*valley-filling*	*do-nothing*
Low	Cheap	BA	VGA	GA
Low	Expensive	BA	GA	GA
Average	Cheap	BA	GA	VGA
Average	Expensive	BA	BA	VGA
High	Cheap	BA	BA	VGA
High	Expensive	VGA	BA	BA

new state. The aim of this reward is to evaluate how much the action taken a_t is suitable for a certain state s_t. Fuzzy logic is used here to evaluate the action taken at a certain state. Fuzzy reasoning is a decision-making model that deals with approximate values rather than exact values. A Fuzzy Inference System (FIS) provides the mapping from the inputs to the outputs, based on a set of fuzzy rule and associated fuzzy Membership Functions (MFs). There are two types of FIS, Mamdani-type FIS and Sugeno-type FIS. Mamdani method is used in this paper because it offers a smoother output. The inputs variables to the fuzzy reward model are the power demand E_t^{total} and the electricity price P_t (referred to as ''states'' in Q-learning) and the outputs variables are the evaluation of *shifting*, *valley-filling* and *do-nothing* (refer to as ''actions'' in Q-learning) as shown in Figure 4.

The MFs for the input variable ''power demand are triangular and are labelled as: Low, Average and High. The universe of discourse of power demand is chosen as [0 6300] (Watt) as shown in Figure 5.

The fuzzy sets of electricity price are defined as ''cheap'' and ''expensive''. The MFs are Gaussian and the universe discourse is [0 0.16] (£/kWh) as shown in Figure 6.

The outputs of the system are the evaluation of the random action which was defined in Q-learning. For each action taken (output), the fuzzy sets are determined as Bad Action (BA), Good Action (GA) and Very Good Action (VGA). The universe of discourse of MFs is defined as [0 100] to evaluate all possible actions with values out of 100 as shown in Figure 7.

Table 4 shows the list of fuzzy rules. Figure 8 illustrates an example of how the FIS evaluates the possible actions for each state. The example shows that the power demand is 5500 W and the electricity price is 0.14 £/kWh which refers to state index 6 according to Table 3. The values of the three actions are 86.5 for *shifting* action, 13.5 for *valley-filling* action and 13.5 for *do-nothing* action. Therefore, if the agent selects *shifting* and will receive a reward of 86.5. Conversely, it will receive a reward of only 13.5 if either *valley-filling* or *do-nothing* action is selected.

V. HOME ENERGY MANAGEMENT ALGORITHM USING Q-LEARNING

Q-learning is considered as an off-policy RL algorithm that seeks to make the best decision at a given state. Off-policy means that the Q-learning function learns from taking random actions without following a current policy. Therefore, a policy is not needed during a training process. The Q-matrix, which has a dimension of [*states* × *actions*], should be initialised to zero (i.e. the Q-value of each state-action pair is signed to zero). Then, the agent will interact with the environment and update each pair in that matrix after each action taken at a certain state using equation (4). In this paper, a random action called ''exploring'' is applied. In this case, a sufficient number of iterations will be required to explore and update the values of $Q(s_t, a_t)$ for all state-action pairs at least once. After convergence of the Q-matrix the optimal Q-values will be obtained.

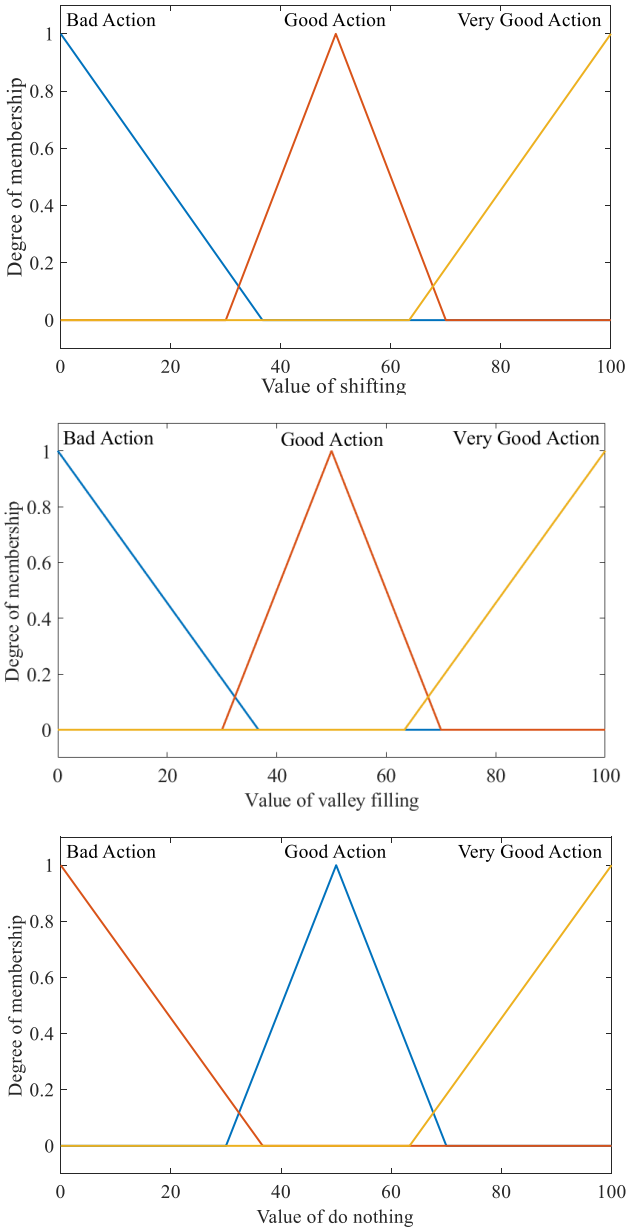

FIGURE 7. Fuzzy sets and MFs of output variables.

The pseudo-code listed in Table 5 (Algorithm 1) illustrates the procedure of the main algorithm of the HEM using Q-learning. Firstly, the numerical rewards are defined using fuzzy logic. The parameters γ and α are set to 0.8 and 0.2 respectively and Q-value matrix entries are initialised to zeros. For each current state, all possible actions are specified, and then an action will be selected randomly. After the selected action is executed, the numerical reward (using fuzzy logic) for that action and the new state will be observed by the agent. The maximum Q-value for the next state should be also determined and then the Q-value of the state-action pair will be updated using equation (4). Finally, the next state will be used as a current state.

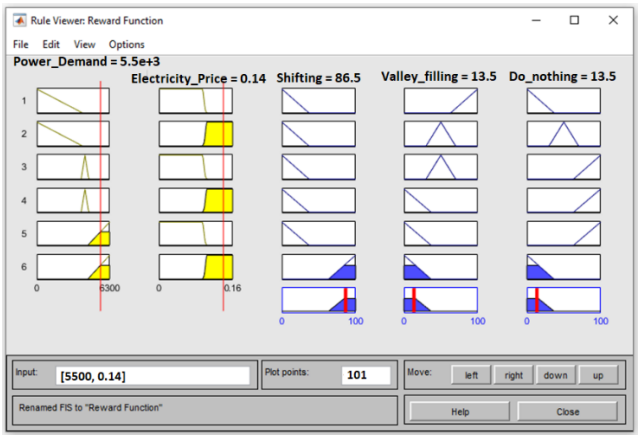

FIGURE 8. Example of FIS process.

TABLE 5. Home energy management using Q-learning algorithm.

Algorithm 1

1. Set $\boldsymbol{\gamma}$,$\boldsymbol{\alpha}$ parameters and environment rewards in matrix as in Table 4.
2. Initialise $\boldsymbol{Q(s_t, a_t)}$,$\forall \boldsymbol{s} \in \boldsymbol{S}, \forall \boldsymbol{a} \in \boldsymbol{A}$.
3. **For** each time step t do
4. Choose a random initial state.
5. **While** hour = 1:24
6. Determine all available actions.
7. Select random action from all possible actions for the current state.
8. Execute the selected action $\boldsymbol{a_t}$, and observe the new state $\boldsymbol{s_{t+1}}$ and numerical reward $\boldsymbol{r(s_t, a_t)}$.
9. Determine the maximum Q-value for next state in Q-matrix.
10. Update the $\boldsymbol{Q(s_t, a_t)}$ using Equation 4.
11. Set the next state as current state.
12. **End while**
13. **End for**

State \ Action	Shifting	Valley filling	Do nothing
1	3.29	2.64	2.80
2	2.80	2.79	3.36
3	2.84	2.79	3.34
4	2.64	2.85	3.19
5	2.82	3.03	3.17
6	2.63	3.38	2.70

Max Q (s_{t+1}, a_{t+1})

State \ Action	Shifting	Valley filling	Do nothing
1	3.29	2.64	2.80
2	2.80	2.79	3.36
3	2.84	2.79	3.34
4	2.64	2.85	3.19
5	2.82	3.03	3.17
6	2.60	3.38	2.70

$Q^*(s_t, a_t)$

FIGURE 9. Simple example of Q-matrix updating.

Figure 9 shows an example of Q-matrix updating. Each row indicates a state and each column indicates an action. Assume that the current state index at time step t is 6 (which represents high power demand and expensive price $[E_{high}^{total}, P_t^{expensive}]$), the random selected action is [Shifting]. Using the fuzzy model, the reward will be obtained as a value of *shifting* action. The next state is observed as 4

TABLE 6. Convergence Q-matrix after 10000 iterations.

State \ Action	Shifting	Valley filling	Do nothing
1	2.57	3.21	2.96
2	2.62	3.01	3.17
3	2.62	3.00	3.27
4	2.92	2.47	3.30
5	2.62	2.87	3.21
6	3.34	2.72	2.76

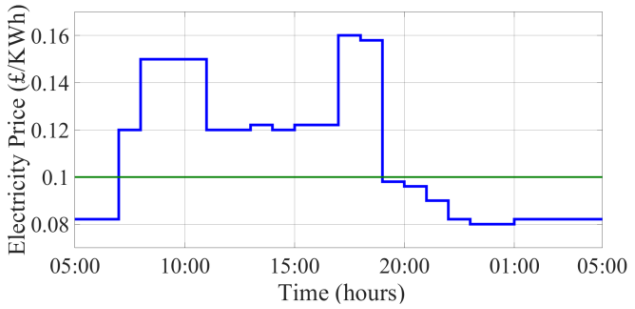

FIGURE 10. Real time price (blue) and average price (green) signals.

(i.e. state $[E^{total}_{average}, P^{expensive}_{t}]$). $\max Q(s_{t+1}, a_{t+1})$ is 3.19 which is found in Q-matrix based on the next state. Using equation (4), the new Q-value for [state: 6, action: 1] is 2.60.

To allow the agent to visit all state-state pairs and learn new knowledge, the training process is set to 1000 iterations. The convergence of the Q-Matrix after execution of this number of iterations is shown in Table 6.

VI. RESULT AND DISCUSSION

Smart meters are used in smart home to receive the price signal from an energy supplier and collect the power data of all household appliances, and then send them to HEMS. Consequently, an optimal decision could be made by HEM system to shift the operating time of the appliance that has the lowest priority during peak demand when required using the convergence Q-Matrix that shown in Table 5, and then turn on the appliance that has been shifted and has the highest priority during off-peak hours. This process works based on the relationship of the real-time price, the consumed power by all household appliances considering load priority and customer comfort preference.

Figure 10 shows the electricity price in £/kWh received from the utility grid.

In Figure 11 is shown the total power demand in Watts of the smart home including all electrical appliances. These two values define the state and are passed to the agent at each time step. Based on the convergence Q-matrix, the action will be selected as the maximum Q-value for that current state.

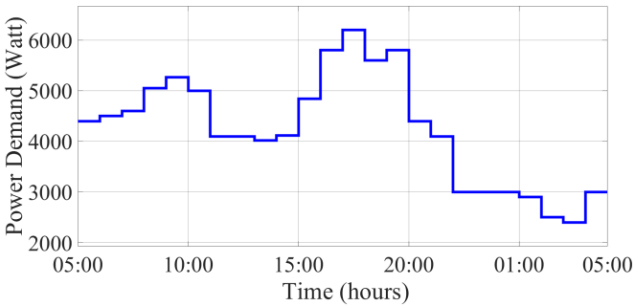

FIGURE 11. Total power demand of the smart home.

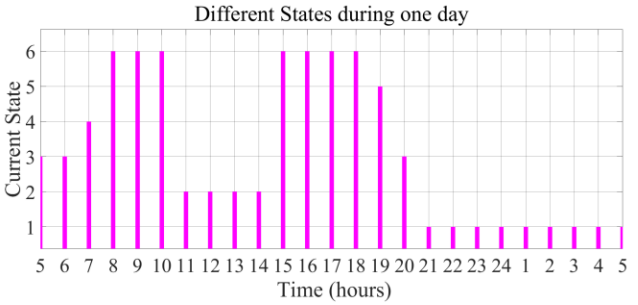

FIGURE 12. All different states based on the price signal and power demand.

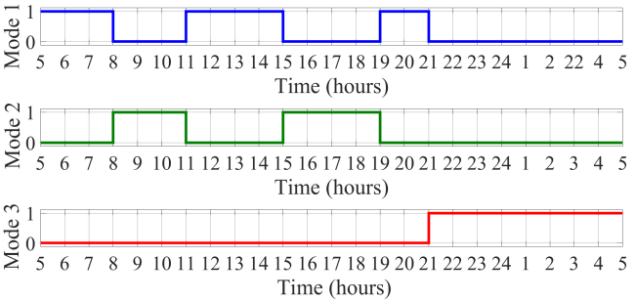

FIGURE 13. Action taken based on the current state and convergence of Q-matrix; Mode 1: Do nothing, Mode 2: Shifting and Mode 3: Valley filling.

Figure 12 shows all different states that are detected based on the different prices and power demand. For example, at 6:00 am the electricity price is low (£0.082) and the power demand is average (4500 W). Thus, the state index is 3. Using Table 5, the maximum value is 3.27 that refers to *do-nothing* action as shown in Figure 13.

At 8:00 am, the energy price is high (£0.15) and the power demand is also high (5000 W). According to Table 3, the current state index is 6. Using Table 5 again, the maximum Q-value is 3.34 which indicates that a *shifting* action should be applied. During night-time, for example at 23:00 pm, the action of *valley-filling* is desirable because the price of electricity is cheap (£0.08) and the power consumption is low (3000 W).

Based on this technique, Figure 14 shows the final power consumption profile of the household appliances over 24 hours. Three periods are identified namely, *shifting* which

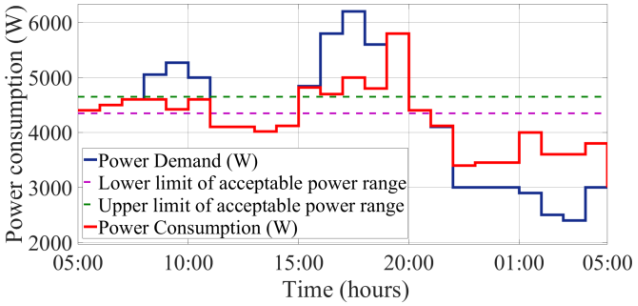

FIGURE 14. Power consumption profile after the implementation of RL algorithm.

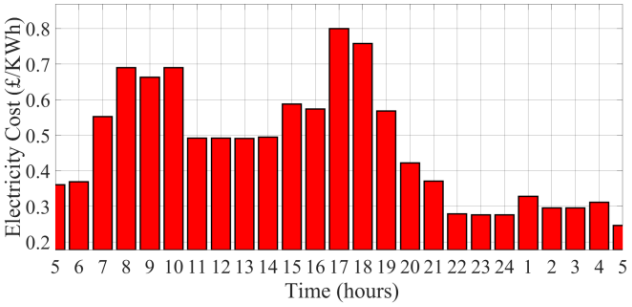

FIGURE 15. Electricity cost without Q-learning.

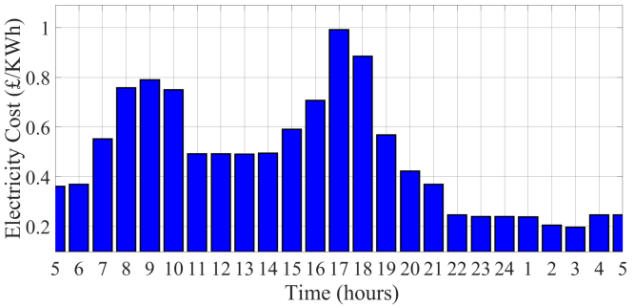

FIGURE 16. Electricity cost with Q-learning.

occurred during [7am-9am] and [15pm-18pm], *valley-filling* occurred during [21pm-5am] and *do-nothing* occurred during [5am-7am], [11am-14pm] and [19pm-20pm].

Figures 15 and 16 show the total electricity cost of all appliances for each hour without and with the Q-value algorithm. The energy cost is reduced during peak demand (when the electricity price is higher). For example, during morning peak demand the energy cost is reduced from £0.8 to £0.7, and from £1.0 to £0.8 during evening peak period. Which demonstrates the effectiveness of the proposed Q-learning-based HEM scheme. To consider the user's comfort, based on the priority of appliances in Table 2, the appliance that has the lowest priority will be shifted during *shifting* mode and that with higher priority. During *valley filling* mode, the shifted appliance that has the highest priority will be turned on.

This study was aimed also to develop a useful user-interface for HEMS algorithms that enables researchers

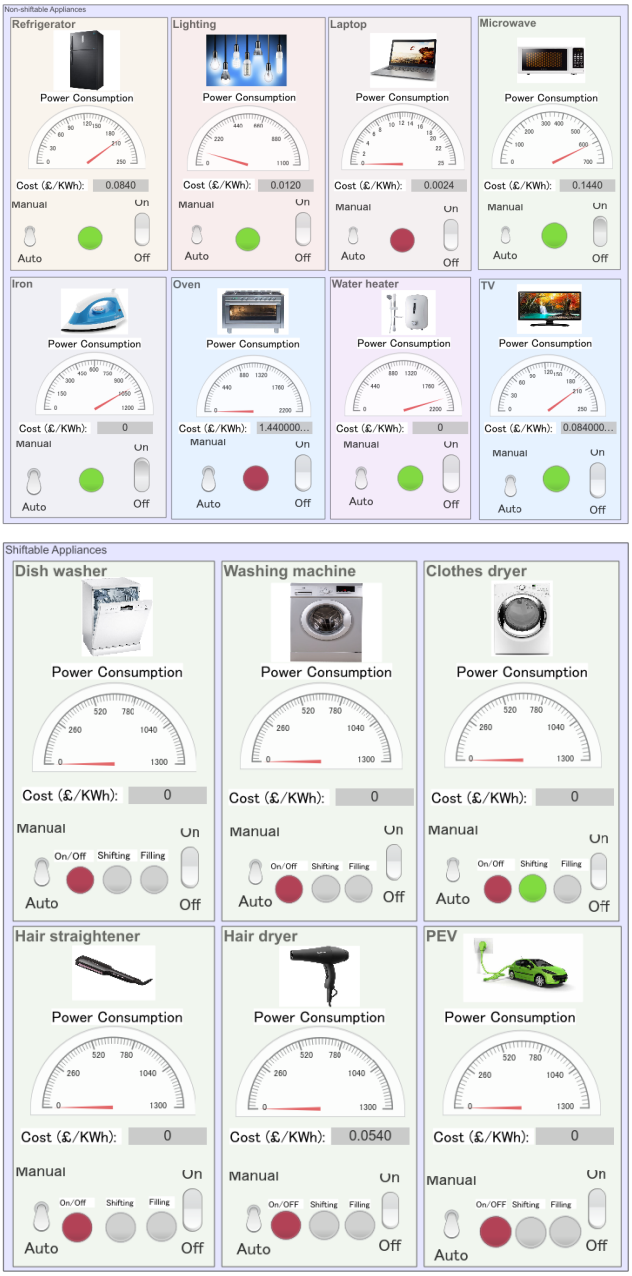

FIGURE 17. User-interface for household appliances implemented using MATLAB/SIMULINK.

and developers to implement and test their proposed control algorithms. The designed user-interface enables the user to control and manage the power consumption and input his/her preference settings. Furthermore, it allows the user to monitor the energy cost for each individual appliance and the total energy cost of all devices. The proposed user-interface provides the user with both auto and manual operation for every appliance as shown in Figure 17. Using auto mode, the system shifts the appliance operating time when required without user's permission and sends an alert signal by lighting the green LED of shifting action. The system turns on automatically the shifted appliance during off-peak by taking into consideration the appliances' priorities, and

then sends a green light signal to indicate the *valley-filling* action.

The system can also operate in manual mode by switching the manual button of the appliance to be controlled manually. This mode is useful when the user wants to override the system by switching on or off each appliance manually.

VII. CONCLUSION

This paper proposed a demand response algorithm to minimise energy utilisation efficiency and electricity bills by shifting load demand, in response to electricity price signal and consumer preferences, from peak periods when the electricity price is high, to off-peak demand when the electricity price is low. In this study, an effective household energy management is developed using Q-learning to deal with the dynamic electricity prices and different power consumption patterns without compromising the users' lifestyle and preferences. The proposed RL-based approach uses a single agent with less number of states and actions to deal with 14 household appliances which in turn makes the implementation much easier, better performance and lower time-consuming comparing to other techniques. Fuzzy reasoning is also used as human thinking to evaluate the random action that the agent could take as a reward function. This helps with avoiding the rules-based technique (crisp values) and obtaining good performance.

The simulation scenarios presented showed that the proposed RL leads to a smooth the power consumption profile and minimises electricity cost by 15% and 18.5% during the morning and evening peak periods respectively, considering user's comfort using priorities for shiftable appliances, user's feedbacks on each action taken and his/her preference settings of the user's interface. Furthermore, the energy costs of two different cases without and with DR were compared to demonstrate how the DR algorithm can contribute to the reduction of electricity cost.

REFERENCES

[1] M. Pollitt and L. Dale, "Restructuring the chinese electricity supply sector-how industrial electricity prices are determined in a liberalized power market: Lessons from Great Britain," Univ. Cambridge Repository, Cambridge, U.K., Tech. Rep. CAM.33977, 2018.

[2] H. Rudnick and C. Velasquez, *Taking Stock of Wholesale Power Markets in Developing Countries: A Literature Review*. Washington, DC, USA: The World Bank, 2018.

[3] M. Shakeri, M. Shayestegan, H. Abunima, S. M. S. Reza, M. Akhtaruzzaman, A. R. M. Alamoud, K. Sopian, and N. Amin, "An intelligent system architecture in home energy management systems (HEMS) for efficient demand response in smart grid," *Energy Buildings*, vol. 138, pp. 154–164, Mar. 2017.

[4] M. Beaudin and H. Zareipour, "Home energy management systems: A review of modelling and complexity," in *Energy Solutions to Combat Global Warming*. Cham, Switzerland: Springer, 2017, pp. 753–793.

[5] P. Stoll, N. Brandt, and L. Nordström, "Including dynamic CO_2 intensity with demand response," *Energy Policy*, vol. 65, pp. 490–500, Feb. 2014.

[6] H. T. Haider, O. H. See, and W. Elmenreich, "A review of residential demand response of smart grid," *Renew. Sustain. Energy Rev.*, vol. 59, pp. 166–178, Jun. 2016.

[7] N. G. Paterakis, O. Erdinç, and J. P. Catalão, "An overview of demand response: Key-elements and international experience," *Renew. Sustain. Energy Rev.*, vol. 69, pp. 871–891, Mar. 2017.

[8] X. Yan, Y. Ozturk, Z. Hu, and Y. Song, "A review on price-driven residential demand response," *Renew. Sustain. Energy Rev.*, vol. 96, pp. 411–419, Nov. 2018.

[9] H. Shareef, M. S. Ahmed, A. Mohamed, and E. Al Hassan, "Review on home energy management system considering demand responses, smart technologies, and intelligent controllers," *IEEE Access*, vol. 6, pp. 24498–24509, 2018.

[10] B. Zhou, W. Li, K. W. Chan, Y. Cao, Y. Kuang, X. Liu, and X. Wang, "Smart home energy management systems: Concept, configurations, and scheduling strategies," *Renew. Sustain. Energy Rev.*, vol. 61, pp. 30–40, Aug. 2016.

[11] G. Huang, J. Yang, and C. Wei, "Cost-effective and comfort-aware electricity scheduling for home energy management system," in *Proc. IEEE Int. Conf. Big Data Cloud Comput. (BDCloud), Social Comput. Netw. (SocialCom), Sustain. Comput. Commun.(SustainCom) (BDCloud-SocialCom-SustainCom)*, Oct. 2016, pp. 453–460.

[12] S. Rajalingam and V. Malathi, "HEM algorithm based smart controller for home power management system," *Energy Buildings*, vol. 131, pp. 184–192, Nov. 2016.

[13] M. Ahmed, "A home energy management algorithm in demand response events for household peak load reduction," *Przeglad Elektrotechniczny*, vol. 1, no. 3, pp. 199–202, Mar. 2017.

[14] A. Ahmad, A. Khan, N. Javaid, H. M. Hussain, W. Abdul, A. Almogren, A. Alamri, and I. Azim Niaz, "An optimized home energy management system with integrated renewable energy and storage resources," *Energies*, vol. 10, no. 4, p. 549, Apr. 2017, doi: 10.3390/en10040549.

[15] K. Ma, T. Yao, J. Yang, and X. Guan, "Residential power scheduling for demand response in smart grid," *Int. J. Electr. Power Energy Syst.*, vol. 78, pp. 320–325, Jun. 2016.

[16] P. Samadi, V. W. S. Wong, and R. Schober, "Load scheduling and power trading in systems with high penetration of renewable energy resources," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1802–1812, Jul. 2016.

[17] R. Lu, S. H. Hong, and M. Yu, "Demand response for home energy management using reinforcement learning and artificial neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6629–6639, Nov. 2019.

[18] C. Zhang, S. R. Kuppannagari, C. Xiong, R. Kannan, and V. K. Prasanna, "A cooperative multi-agent deep reinforcement learning framework for real-time residential load scheduling," in *Proc. Int. Conf. Internet Things Desigh Implement. (IoTDI)*, 2019, pp. 59–69.

[19] S. Lee and D. H. Choi, "Reinforcement learning-based energy management of smart home with rooftop solar photovoltaic system," *Energy Storage Syst. Home Appl. Sensors*, vol. 19, no. 18, p. 3937, 2019.

[20] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Appl. Energy*, vol. 235, pp. 1072–1089, Feb. 2019.

[21] S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," *Energies*, vol. 11, no. 8, p. 2010, 2018.

[22] N. Chauhan, N. Choudhary, and K. George, "A comparison of reinforcement learning based approaches to appliance scheduling," in *Proc. 2nd Int. Conf. Contemp. Comput. Inform. (ICI)*, Dec. 2016, pp. 253–258.

[23] M. Ul Hassan, M. H. Rehmani, R. Kotagiri, J. Zhang, and J. Chen, "Differential privacy for renewable energy resources based smart metering," *J. Parallel Distrib. Comput.*, vol. 131, pp. 69–80, Sep. 2019.

[24] Z. Guan, G. Si, X. Zhang, L. Wu, N. Guizani, X. Du, and Y. Ma, "Privacy-preserving and efficient aggregation based on blockchain for power grid communications in smart communities," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 82–88, Jul. 2018.

[25] M. S. Ahmed, A. Mohamed, T. Khatib, H. Shareef, R. Z. Homod, and J. A. Ali, "Real time optimal schedule controller for home energy management system using new binary backtracking search algorithm," *Energy Buildings*, vol. 138, pp. 215–227, Mar. 2017.

[26] Z. Zhao, K. Agbossou, and A. Cardenas, "Connectivity for home energy management applications," in *Proc. IEEE PES Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Oct. 2016, pp. 2175–2180.

[27] J. M. G. Lopez, E. Pouresmaeil, C. A. Canizares, K. Bhattacharya, A. Mosaddegh, and B. V. Solanki, "Smart residential load simulator for energy management in smart grids," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1443–1452, Feb. 2019.

[28] Z. Wan, H. Li, and H. He, "Residential energy management with deep reinforcement learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7.
[29] R. Lu, S. H. Hong, and X. Zhang, "A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach," *Appl. Energy*, vol. 220, pp. 220–230, Jun. 2018.
[30] F. Alfaverh, M. Denai, and K. Alfaverh, "Demand-response based energy advisor for household energy management," in *Proc. 3rd World Conf. Smart Trends Syst. Secur. Sustainablity*, Jul. 2019, pp. 153–157.

FAYIZ ALFAVERH received the B.Sc. degree in electrical power engineering from Yarmouk University, Irbid, Jordan, and the M.Sc. degree in power electronics and control engineering from the University of Hertfordshire, Hatfield, U.K., where he is currently pursuing the Ph.D. degree. His interests are in demand response management and control strategies for integrated smart electricity networks, management of smart homes and dynamic scheduling, and the optimization and control of future smart grids.

M. DENAÏ received the degree in electrical engineering from the University of Science and Technology of Algiers and Ecole Nationale Polythechnique of Algiers, Algeria, and the Ph.D. degree in control engineering from the University of Sheffield, U.K.

He worked for the University of Science and Technology of Oran, Algeria, until 2004, and the University of Sheffield, from 2004 to 2010. From 2010 to 2014, he worked for the University of Teesside, U.K. He has been with the University of Hertfordshire, U.K., since 2014. His main fields of expertise are in modeling, optimization, and control of engineering and life science (biological and biomedical) systems. His current research interests in energy include intelligent control design and computational intelligence applications to efficiency optimization in renewable energy systems with particular focus in the management of smart homes and dynamic scheduling, optimization and control of future smart grids, condition monitoring and asset management in electric power networks; energy storage systems integration into the grid; smart meter data analytics using machine learning techniques for efficient energy management; and electric vehicles integration into the distribution grid and V2G/G2V management.

YICHUANG SUN (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from Dalian Maritime University, Dalian, China, in 1982 and 1985, respectively, and the Ph.D. degree from the University of York, York, U.K., in 1996, all in communications and electronics engineering.

He is currently a Professor of communications and electronics, the Head of the Communications and Intelligent Systems Research Group, and the Head of the Electronic, Communication and Electrical Engineering Division, School of Engineering and Computer Science, University of Hertfordshire, U.K. He has published more than 330 articles and contributed 10 chapters in edited books. He has also published four text and research books *Continuous-Time Active Filter Design* (CRC Press, USA, 1999), *Design of High Frequency Integrated Analogue Filters* (IEE Press, U.K., 2002), *Wireless Communication Circuits and Systems* (IET Press, 2004), and *Test and Diagnosis of Analogue, Mixed-Signal and RF Integrated Circuits—the Systems on Chip Approach* (IET Press, 2008). His research interests are in the areas of wireless and mobile communications, RF and analogue circuits, microelectronic devices and systems, machine learning, and deep learning.

Prof. Sun was a Guest Editor of eight IEEE and IEE/IET journal special issues High-frequency Integrated Analogue Filters in *IEE Proceedings of Circuits, Devices and Systems*, in 2000, RF Circuits and Systems for Wireless Communications in *IEE Proceedings of Circuits, Devices and Systems*, in 2002, Analogue and Mixed-Signal Test for Systems on Chip in *IEE Proceeding of Circuits, Devices and Systems*, in 2004, MIMO Wireless and Mobile Communications in *IEE Proceedings of Communications*, in 2006, Advanced Signal Processing for Wireless and Mobile Communications in *IET Signal Processing* (2009), Cooperative Wireless and Mobile Communications in *IET Communications* (2013), Software-Defined Radio Transceivers and Circuits for 5G Wireless Communications in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II, in 2016, and the 2016 IEEE International Symposium on Circuits and Systems in the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I, in 2016. He was a Series Editor of IEE Circuits, Devices and Systems Book Series from 2003 to 2008. He has been an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: Regular Papers from 2010 to 2011, from 2016 to 2017, and from 2018 to 2019. He is also an Editor of *ETRI Journal*, the *Journal of Semiconductors*, and the *Journal of Sensor and Actuator Networks*. He has also been widely involved in various IEEE technical committee and international conference activities.