# HOME ENERGY MANAGEMENT SYSTEMS AND METHODS

## TECHNICAL FIELD

[0001] The present disclosure relates to energy management systems for smart homes.

## BACKGROUND

[0002] Advanced metering infrastructure (AMI) can collect user electricity consumption with time stamps. The worldwide roll out of AMI opens up opportunities to reduce electricity cost and improve energy efficiency through leveraging the use of big data. Home energy management systems (HEMS), which monitor and manage electricity or thermal energy generation, storage and consumption in smart houses, are one application which utilizes data collected by AMI. HEMS is considered as a recommendation system that helps user gain profit in price-based demand response (DR) program.

[0003] Previously, the production and consumption of electricity and thermal energy in a smart home were independent of each other, electricity was obtained from the grid and consumed by electrical loads, while thermal energy was directly produced by gas heat pump and consumed by thermal loads. However, In recent years, battery energy storage system (BESS), cogeneration micro combined heat and power (mCHP) systems and thermal energy storage (TES) system are applied at the building level to achieve higher energy utilization efficiencies, at the same time, the interact of thermal energy and electricity balance also brings a potential to save total energy cost in a dynamic electricity price policy, the electricity and thermal energy balance are becoming highly coupled and the control variable increase exponentially. Thus, HEMS for a multi-energy smart home is a more challenging problem to solve.

[0004] In the early stage, many researchers apply model-based control methods on energy management. Optimization method such as dynamic programming, robust optimization and stochastic programming are widely applied in this field to solve different objective function that considering electricity cost of different types of

appliances, user's satisfaction level, daily peak-average ratio (PAR). However, these approaches only play a good role when all environment information is known, which will require the knowledge of all accurate statistical distribution of environmental parameters, which is not realistic as the estimation is not accurate and there are always prediction errors in forecast data such as the electricity price, energy consumption and PV generation. Thus, traditional optimization methods are not suitable to solve these variables under uncertainties. In addition, as the numbers of appliances get larger, the computational cost will increase exponentially.

[0005] In recent years, with the deepening of artificial intelligence research and the continuous improvement of computer computing power, reinforcement learning (RL) has been widely applied in more and more fields due to its excellent decision-making ability. Reinforcement learning is a model-free method which does not require any initial information, its environment is considered as a black box that receives an action signal in a discrete-time step from its agent and gives a reward to the agent. The agent is rewarded and the environment changes to the new state according to the previous state and the action taken in this time step. RL attempts to establish a mapping between agent action and state to maximize the accumulated rewards in an unsupervised manner.

[0006] Up to now, some research work has applied RL to user-side scenarios such as smart home energy management systems and EV charging management. For example, various Q-learning approaches have been proposed. These approaches based on Q learning algorithms strictly based on assumption that all appliance work on discrete power levels, which is unrealistic in real applications. In order to make HEMS work in a continuous range, a variety of different DRL methods are applied in this field. However, most research are still focusing on the establishment of models and agents to plan for one or a few electrical appliances, which will ignore the cooperation and competition between components. Besides, very little research considers the planning of smart homes in at multi-energy level.

SUMMARY

[0007] According to a first aspect of the present disclosure, a home energy management system for a smart home is provided. The smart home has an electrical grid connection, a renewable energy source which provides a time varying electrical power output, a battery energy storage system, and a plurality of electrical loads. The home energy management system comprises: a processor which implements a reinforcement learning agent trained to select control actions for the battery energy storage system and the plurality of electrical loads which minimize an objective function dependent on an energy cost parameter, a comfort metric of the smart home and a degradation metric of the battery energy storage system; and a controller configured to control the battery energy storage system and the plurality of electrical loads according to the control actions selected by the reinforcement learning agent.

[0008] The present disclosure provides a Deep Deterministic Policy Gradient (DDPG) based multi-energy home energy management system (HEMS). The HEMS is proposed to schedule the key components in a multi-energy smart home. The proposed method works on continuous control action space and gives solutions to the utilization of rooftop PV and ESS for both electricity and thermal energy, long-term battery degradation costs of battery ESS are considered. The present disclosure provides the following advantages. Energy cost minimization problem in a multi-energy smart home which includes BESS, electrical loads, EV, PV, thermal load, TES, EHP, CCHP, gas boiler is investigated in this paper. A discrete-time MDP process is modelled to represent the operation of the multi-energy smart home under uncertainties (PV and load). User's dissatisfaction of load shifting in power level and time scale is considered.

[0009] A model-free and data-driven approach based on deep deterministic policy gradient (DDPG) method is proposed to determine the control action in continuous workspace for previous mentioned multi-energy smart home, by adding the normalization layer and denormalization layer in basic DDPG neural networks structure, the developed method manage to work on a constraint action space. Case studies on a real-world scenario with open-source electricity price data demonstrate that the proposed method significantly lowers the daily energy cost considering the user's satisfactory level.

[0010] In an embodiment, the reinforcement learning agent is a deep deterministic policy gradient agent.

[0011] In an embodiment, the plurality of electrical loads comprises a power shiftable load.

[0012] In an embodiment, wherein the plurality of electrical loads comprises a time shiftable load.

[0013] In an embodiment, the plurality of electrical loads comprises an electrical vehicle charging load. The electrical vehicle charging load may have a specified charging period in which an electrical vehicle must be charged.

[0014] In an embodiment, the processor further implements an energy cost forecast module configured to predict energy cost. The energy cost forecast module may comprise an artificial neural network trained to predict energy cost.

[0015] In an embodiment, the smart home has a gas input connection and further comprises a gas boiler configured to generate thermal energy from gas, and wherein the reinforcement learning agent is further trained to select control actions for the gas boiler to minimize the objective function.

[0016] According to a second aspect of the present disclosure, an energy management method for a smart home is provided. The smart home has an electrical grid connection, a renewable energy source which provides a time varying electrical power output, a battery energy storage system, and a plurality of electrical loads. The energy management method comprises: receiving environmental state indications indicating an environmental state of the smart home; selecting, by a reinforcement learning agent, control actions for the battery energy storage system and the plurality of electrical loads which minimize an objective function dependent on an energy cost parameter, a comfort metric of the smart home and a degradation metric of the battery energy storage system; and controlling the battery energy storage system and the plurality of electrical loads according to the control actions selected by the reinforcement learning agent.

[0017] In an embodiment, the reinforcement learning agent is a deep deterministic policy gradient agent.

[0018] In an embodiment, the plurality of electrical loads comprises a power shiftable load.

[0019] In an embodiment, the plurality of electrical loads comprises a time shiftable load.

[0020] In an embodiment, the plurality of electrical loads comprises an electrical vehicle charging load. The electrical vehicle charging load may have a specified charging period in which an electrical vehicle must be charged.

[0021] In an embodiment, the method further comprises predicting future energy costs. The future energy costs may be predicted using an artificial neural network trained to predict energy cost.

[0022] In an embodiment, the smart home has a gas input connection and further comprises a gas boiler configured to generate thermal energy from gas, and the energy management method further comprises selecting control actions for the gas boiler to minimize the objective function.

[0023] According to a third aspect of the present disclosure a computer readable medium storing processor executable instructions which when executed on a processor cause the processor to carry out a method as set out above is provided.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0024] In the following, embodiments of the present invention will be described as non-limiting examples with reference to the accompanying drawings in which:

[0025] FIG.1 shows the structure of a smart home which is managed by a home energy management system according to an embodiment of the present invention;

[0026] FIG.2 is a block diagram showing a home energy management system according to an embodiment of the present invention;

[0027] FIG.3 is a block diagram showing a deep deterministic policy gradient actor critic model used in embodiments of the present invention;

[0028] FIG.4 is a flow chart showing an energy management method for a smart home according to an embodiment of the present invention;

[0029] FIG.5 shows a comparison of predicted electricity prices calculated by an electricity cost forecast module of an embodiment of the present invention with actual electricity prices;

[0030] FIG.6 shows the rewards of each training iteration in an example embodiment of a home energy management system according to an embodiment of the present invention;

[0031] FIG.7 is a graph showing electrical balance of each hour for a smart home implementing a home energy management system according to an embodiment of the present invention;

[0032] FIG.8 is a graph showing thermal energy balance of each hour for a smart home implementing a home energy management system according to an embodiment of the present invention;

[0033] FIG.9 is a graph showing an example of energy usage in a smart home without a home energy management system;

[0034] FIG.10 is a graph showing an example of energy usage in a smart home with a home energy management system according to an embodiment of the present invention;

[0035] FIG.11 is a graph showing EV charging scheduling given by a home energy management system according to an embodiment of the present invention; and

[0036] FIG.12 is a graph showing state of charge for a battery energy storage system and thermal energy storage system in a smart home using an energy management system according to an embodiment of the present invention.

DETAILED DESCRIPTION

[0037] In the present disclosure, a smart home with a variety of different types of loads, electrical and thermal energy storage systems, renewable energy sources (photovoltaics), electrical vehicle, cogeneration, electrical heat pump and gas boiler is considered. Under the dynamic electricity price tariff (take day ahead electricity price as an example), a deep reinforcement learning-based energy management system is developed to schedule controllable components on a day-ahead basis, so as to achieve the minimization of the daily energy cost. The mathematical models of the key components, including the various operational constraints and the objective function for different types of appliances, energy storage system, are described.

[0038] FIG.1 shows the structure of a smart home which is managed by a home energy management system according to an embodiment of the present invention.

[0039] As shown in FIG.1, the smart home 100 is connected to the electricity grid 110 which supplies electrical power and to a gas input 120 which supplies gas. The smart home 100 is provided with photovoltaic (PV) system 112 which generates renewable electricity, a battery energy storage system (ESS) 114 which can store electrical energy, an electrical vehicle (EV) charging system 116 and electrical loads 118. As will be described in more detail below, the electrical loads may comprise power shiftable loads, time shiftable loads and non-shiftable loads. The gas input 120 is connected to a cogeneration system 122 which converts heat generated from burning gas into electricity and a gas boiler 124. Heat may be stored in a thermal tank 126 and is consumed by thermal loads 128. An electrical boiler 130 converts electrical energy into heat if required. The daily energy cost includes the electricity and gas bought from

the utility, and in this disclosure, it is assumed that the electricity prices change hourly, while unit natural gas price is fixed.

[0040] It should be noted that while FIG.1 shows a photovoltaic system, it is envisaged that embodiments of the present invention may operate with other renewable energy sources such as wind turbines. Generally, renewable energy sources will have an electrical power output which varies over time depending on, for example, current weather conditions. Further, it will be appreciated that embodiments of the present invention may operate with smart homes which include a subset of the devices shown in FIG.1.

[0041] The components of the smart home and the objectives of the smart home energy management system will now be described in more detail.

Battery Energy Storage System (BESS)

[0042] With the BESS costs rapidly dropping with the development of technology and are predicted to drop below 100$/kWh within the next decade. As a result, the profitability of adopting ESS will gradually increase.

[0043] The hourly operation of BESS is given in equation (1), where $SOC_t$ represents the state of charging of BESS in hour *t*, $E_{t,C}^{BESS}$is the electricity charging into BESS in hour *t*. $E_{t,D}^{BESS}$is the electricity discharged by BESS in hour *t*.

$$0 \leq SOC_{t+1} = SOC_t + E_{t,C}^{BESS} - E_{t,D}^{BESS} \leq C^{BESS} \quad (1)$$

$$0 < E_{t,C}^{BESS} < E_{max\,C}^{BESS} \quad (2)$$

$$0 < E_{t,D}^{BESS} < E_{max\,D}^{BESS} \quad (3)$$

$$E_{t,C}^{BESS} * E_{t,D}^{BESS} = 0 \quad (4)$$

[0044] Equations (2) and (3) give the limit of hourly charging or discharging electricity, equation (4) constrains ESS cannot charging and discharging in the same hour.

[0045] To estimate the battery long-term degradation cost, considering that the number of charge and discharge times of BESS in a day is quite small compared to its life cycle, the change in battery capacity due to degradation within a day can be ignored. The battery degradation cost of each charging or discharging process can be estimated using battery change cost, current capacity, battery degradation curve and the charging or discharging electricity. Equation (5) is the relationship between the number of life cycles and degree and degradation (DOD) of Ni-Cd batteries, where $a, b, c$ are curve-fitting coefficients. The battery lifetime has the best fitting in the following expression with $DOD = d_B$. Equation (6) is the estimation of the long-term battery degradation cost.

$$L_B(d_B) = a \times {d_B}^{-b} \times e^{-cd_B} \tag{5}$$

$$C_{BDC}\big(t, d_B(\Delta t)\big) = C_{BAC}\big(t, d_B(\Delta t)\big)P_B(t) = \frac{C_B P_B(t)\Delta t}{2L_B\big(d_B(\Delta t)\big)E_{BA}(t)d_B(\Delta t)\eta_{Bc}\eta_{Bd}} \tag{6}$$

Thermal Energy Storage System (TES)

[0046] The operation principle of TES is same as on the general principle of BESS and is also governed by the flow of heat energy which is limited by the energy storage limits. The hourly operation and constraints of TES is given in equation (7)~(10):

$$0 \le E_{t+1}^{TES} = E_t^{TES} + H_{t,C}^{TES} - H_{t,D}^{TES} \le C^{TES} \tag{7}$$

$$0 < H_{t,C}^{TS} < H_{max\,C}^{TS} \tag{8}$$

$$0 < H_{t,D}^{TS} < H_{max\,D}^{TS} \tag{9}$$

$$H_{t,C}^{TS} * H_{t,D}^{TS} = 0 \tag{10}$$

Combined heat and power system

[0047] Combined heat and power (CHP) system, also known as cogeneration, has recently been identified as a potential solution to achieve higher energy utilization

efficiency[18]. CHP simultaneous production of two or more forms of energy from a single fuel source.

[0048] At home level, micro combined heat and power (mCHP) can be used in homes with the ratting of 5 kW or below. In recent year some mCHP system are commercially available for building applications, manufactured by the Panasonic, GE, Toshiba, BlueGen, among other companies. Institutional efforts include the Japanese Ene-Farm introduced by the Panasonic and Tokyo Gas. Typically, mCHP consumes natural gas and generate electricity, waste heat is captured with a heat recovery unit to as a thermal energy source.

[0049] Equation (11) to (13) represent electricity and thermal energy that generated by mCHP system hourly, where $\eta_{e,i}$ and $\eta_{th,i}$ are the efficiency, $\beta$ is the conversion factor.

$$E_t^{CHP} = F_t^{CHP} * \eta_{e,i} * \beta \quad (11)$$

$$H_t^{CHP} = F_t^{CHP} * \eta_{th,i} * \beta \quad (12)$$

$$F_{min}^{CHP} \leq F_t^{CHP} \leq F_{max}^{CHP} \quad (13)$$

Electrical Heat Pump (EHP)

[0050] EHP consumes electricity to generate heat for thermal load consumption and thermal storage charging. the thermal energy generated by the EHP and the constraint are given in equation (14) and (15), where $\eta_{EHP}$ is the efficiency of EHP.

$$H_t^{EHP} = E_t^{EHP} * \eta_{EHP} \quad (14)$$

$$E_{min}^{EHP} \leq E_t^{EHP} \leq E_{max}^{EHP} \quad (15)$$

Gas Boiler

[0051] Gas boilers (GB) consume natural gas to generate thermal energy, the thermal energy generated in each time slot is proportional to the gas consumed in the time slot, as equation (16), where $\eta_{GB}$ is the effiency of GB.

$$H_t^{GB} = F_t^{GB} * \eta_{GB} * \beta \quad (16)$$

[0052] Different Type of Loads

Power Shiftable Load

[0053] Power shiftable loads (PSL) can be operated with a flexible power level range, such as the air conditioners. PSL power level can be scheduled to respond to price changes. PSL operating range is given in equation (17):

$$E_{min,n}^{PSL} < E_{t,n}^{PSL} < E_{max,n}^{PSL} \quad (17)$$

[0054] The change of PSL consumption will bring the user's dissatisfactory, this dissatisfaction cost is modelled as in equation (18), where $e_{t,n}^{PSL}$ , user's original consumption. $\alpha^{PSL}$ is a device and user-dependent dissatisfaction cost parameter: a device with a greater $\alpha^{PSL}$ prefers to consume more energy to improve the satisfaction level and vice versa.

$$P_{t,n}^{PSL} = \alpha^{PSL}(e_{t,n}^{PSL} - E_{t,n}^{PSL})^2 \quad (18)$$

Daily Operation Load

[0055] Daily operation load (DOL) also can be operated with a flexible power level range like PSL but is expected to consume no less than a certain amount of electricity within a day to support the user's daily work such as water heater and personal computer. Regarding this, constraint (19) need to be added in addition to the above mentioned:

$$E_{min,n}^{PSL} < E_{t,n}^{PSL} < E_{max,n}^{PSL} \quad (19)$$

$$\sum_{t=0}^{23} E_{t,n}^{DL} \geq E_{min} \quad (20)$$

Time Shiftable Load

[0056] Time shiftable loads are designated to complete certain tasks during the day, such as washing machines, dishwashers, etc. This type of loads' operating time can be shifted to off-peak hours to reduce cost. TSL is not shiftable on power level.

[0057] Hourly consumption of time shiftable load is either 0 or a determined amount as shown in (21):

$$E_{t,n}^{TSL} \in \left\{0, E_{st,n}^{TSL}, E_{st+1,n}^{TSL}, \dots E_{st+k,n}^{TSL}\right\} \qquad (21)$$

[0058] The subscript represents the energy consumption from time shiftable load is turned on to it work done.

[0059] The shifting of TSL will also bring user dissatisfaction, equation (22) is the cost function where $\alpha^{TSL}$ is a user and device-dependent dissatisfaction cost parameter: a device with a greater $\alpha^{TSL}$ prefers working at pre-set hours without load shifting.

$$P_n^{TSL} = \alpha^{TSL}(nt - st)^2 \sum_{p=0}^{k} E_{st+p,n}^{TSL} \qquad (22)$$

Non-Shiftable Load

[0060] Non-shiftable load, or base load, refers to electrical appliances that cannot be controlled and must meet certain operating conditions, such as refrigerators and safety warning systems. The consumption is given in (23):

$$E_{t,n}^{NSL} = e_{t,n}^{NSL} \qquad (23)$$

Electrical Vehicle (EV)

[0061] In this disclosure, an ideal EV charging planning model target to minimize the charging cost in dynamic price policy and ensure users' daily energy needs for electric vehicles is proposed. HEMS needs to schedule the EV to charge in a safe charging period between user departure time and arrival time.

[0062] The operation and constraints of EV charging planning are represented as (24) to (28):

$$SOC_t^{EV} = \begin{cases} SOC_{t-1}^{EV} + E_{t-1}^{EV}, \text{if } t \le t_d \text{ or } t > t_a \\ SOC_a, \text{if } t = t_d \end{cases} \quad (24)$$

$$0 \le SOC_t^{EV} \le C^{EV} \quad (25)$$

$$0 \le E_t^{EV} \le E_{max}^{EV} \quad (26)$$

$$SOC_d^{EV} \ge E_{target} \quad (27)$$

$$E_t^{EV} = 0, \text{if } t \in [t_d, t_a) \quad (28)$$

Electricity balance and thermal balance

[0063] In every time slot, both thermal energy and electricity consumption is equal to the input.

[0064] Electricity balance is represented as the equation (29), where BESS charging efficiency is $\eta_{Bc}$ and discharging efficiency is $\eta_{Bd}$.

$$E_t^{Grid} + E_t^{PV} + E_{t,D}^{BESS} * \eta_{Bd} + E_t^{CHP} = E_t^{EV} + E_t^{EHP} + \sum_{type} \sum_{n} E_{t,n}^{type} + \frac{E_{t,C}^{BESS}}{\eta_{Bc}} \quad (29)$$

[0065] Heat balance is represented as the equation (30), where TES charging efficiency is $\eta_{Tc}$ and discharging efficiency is $\eta_{Td}$.

$$H_t^{CHP} + H_t^{GP} + H_{t,D}^{TES} * \eta_{Td} + H_t^{EHP} = \sum_{type} \sum_{n} H_{t,n}^{type} + \frac{H_{t,C}^{TES}}{\eta_{Tc}} \quad (30)$$

Objective Function

[0066] The objective of HEMS is to minimize the total cost function include total energy cost, user dissatisfactory cost and battery long-term degradation cost.

[0067] The electricity bought from grid $E_t^{Grid}$ can be obtained from (29). The gas consumed by gas boiler $F_t^{GB}$ can be obtained from (16) and (30).

[0068] Thus, the cost function to be minimized include the objective of the proposed EMS is given as (31):

$$C = \sum_{t=0}^{t=23} ( E_t^{Grid} * p_t + (F_t^{GB} + F_t^{CHP}) * p^{gas} + C_{BDC}(t) + \sum_{type} \sum_{n} P_{t,n}^{type} \quad (31)$$

[0069] FIG.2 is a block diagram showing a home energy management system according to an embodiment of the present invention. The home energy management system 200 is a computer system with memory that stores computer program modules which implement home energy management methods according to embodiments of the present invention.

[0070] The home energy management system 200 comprises a processor 210, a working memory 212, an interface 214, a controller 216, program storage 220 and data storage 230. The processor 210 may be implemented as one or more central processing unit (CPU) chips. The program storage 220 is a non-volatile storage device such as a hard disk drive which stores computer program modules. The computer program modules are loaded into the working memory 212 for execution by the processor 210. The interface 214 allows data, such as metering data and environmental data to be received by home energy management system 200. The interface 214 may be a wireless network interface such as a Wi-Fi or Bluetooth interface, alternatively it may be a wired interface. The controller 216 allows the home energy management system 200 to control the various components of the smart home 100 shown in FIG.1.

[0071] The program storage 220 stores a reinforcement learning agent module 222, a reinforcement learning training module 224, an electricity cost forecast module 226,

and a control module 228. The computer program modules cause the processor 210 to execute various home energy management processing which is described in more detail below. The program storage 220 may be referred to in some contexts as computer readable storage media and/or non-transitory computer readable media. As depicted in FIG.2, the computer program modules are distinct modules which perform respective functions implemented by the home energy management system 200. It will be appreciated that the boundaries between these modules are exemplary only, and that alternative embodiments may merge modules or impose an alternative decomposition of functionality of modules. For example, the modules discussed herein may be decomposed into sub-modules to be executed as multiple computer processes, and, optionally, on multiple computers. Moreover, alternative embodiments may combine multiple instances of a particular module or sub-module. It will also be appreciated that, while a software implementation of the computer program modules is described herein, these may alternatively be implemented as one or more hardware modules (such as field-programmable gate array(s) or application-specific integrated circuit(s)) comprising circuitry which implements equivalent functionality to that implemented in software.

[0072] The data storage 230 stores an electricity cost forecast model 232 and actor critic model 234. The electricity cost forecast model 232 is an artificial neural network (ANN) which is trained to predict day ahead electricity prices. In an embodiment, a four layer back propagation neural network (BPNN) is used as the electricity cost forecast model 232. Data will be input into the neural network after feature extraction and normalization, hourly electricity price will be used as target output in training process. Features extracted are given in the table below.

| Feature index | Description |
|---|---|
| 1 | Hour index (1~24) |
| 2 | Day of week (1~7) |
| 3 | Day ahead hourly demand of this hour |
| 4 | Electricity price of h-24 |
| 5 | Electricity price of h-48 |

| | |
|---|---|
| 6 | Electricity price of h-72 |
| 7 | Hourly demand of h-24 |
| 8 | Hourly demand of h-48 |
| 9 | Hourly demand of h-72 |

[0073] The actor critic model 234 is a discrete Markov decision process (MDP) model which is used to determine how to control the smart home based on sensed environment states. To perform the optimal action-making, the problem formulated in this disclosure is considered as a discrete Markov Decision Process (MDP), which exhibits the Markov property that the state transitions are dependent only on the current state and the current action taken and, thus are independent of all previous environmental states and agent. The actor critic model 234 is described below with reference to FIG.3.

[0074] FIG.3 is a block diagram showing a deep deterministic policy gradient actor critic model used in embodiments of the present invention. As shown in FIG.3, the deep deterministic policy gradient actor critic model 300 comprises an agent 302 which makes decisions based on an environment 304. The model 300 further comprises an actor network 312 which is a neural network which is trained to output a decision of an action a, according to the state s, of the environment 304 and a critic network 314 which is a neural network trained to determine whether the decision is good or not. The output of the critic network 314 is fed back to the actor network 312 via a policy gradient 316.

[0075] The deep deterministic policy gradient actor critic model 300 makes use of off-policy learning. A replay buffer 306 stores a collection of experiences observed by the agent 302. To improve the stability of the model off-policy target networks are trained from the data stored in the replay buffer 306. As shown in FIG.3, a target actor network 322 are an target critic network 324 are trained from the states and actions stored in the replay buffer 306.

[0076] A bellman loss function 318 is used to learn the Q-function and thereby optimize the actions for a given state.

[0077] An MDP model includes four essential elements: state (*s*), action (*a*), state transition probability (*p*), and reward (*r*), which are modelled as follows:

State (s) :

[0078] State for time slot *t* is given in (32):

$$s_t = \{t, p_t, p_{t+1}, \dots, p_{t+5}, E_t^{PV}, E_{t+1}^{PV}, \dots, E_{t+5}^{PV}, \quad \frac{SOC_t^{BESS}}{C^{ESS}}, \frac{SOC_t^{TES}}{C^{TES}}, \frac{SOC_t^{EV}}{C^{EV}}\} \quad (32)$$

[0079] Where $p_t, p_{t+1}, \dots, p_{t+5}$ is the price for current time slot and following time slots, $E_t^{PV}, E_{t+1}^{PV}, \dots, E_{t+5}^{PV}$ is the PV output for current time slot and following time slots, $\frac{SOC_t^{BESS}}{C^{ESS}}, \frac{SOC_t^{TES}}{C^{TS}}, \frac{SOC_t^{EV}}{C^{EV}}$ are the normalized state of charging (SOC) of BESS, TES and EV correspondingly.

action (a):

[0080] The action for time slot *t* includes all the energy consumption of electrical load, thermal load, EV and electrical heat pump, the charging or discharging energy of BESS and TES, and the fuel consumption of mCHP, which is given in (33):

$$a_t = \{E_{t,n}^{EL}, H_{t,n}^{TL}, E_t^{EV}, E_t^{EHP}, E_{t,C}^{BESS}, E_{t,D}^{BESS}, H_{t,C}^{TES}, H_{t,D}^{TES}, F_t^{CHP}\} \quad (33)$$

Calculation of reward (r):

[0081] The object function is given in the last subsection, as it is a minimize problem, the reward of the given MDP problem is given in (34):

$$r_t = -C = -\Big(\sum_{t=0}^{t=23} \big(E_t^{Grid} * p_t + (F_t^{GB} + F_t^{CHP}) * p^{gas} + C_{BDC}(t) + \sum_{type}\sum_{n} P_{t,n}^{type}\big)\Big) \quad (34)$$

[0082] Reinforcement learning (RL) is an area of machine learning concerned with how intelligent agents ought to take ideal actions in a stochastic environment which is typically stated in the form of a Markov decision process (MDP), in order to maximize the notion of cumulative reward.

[0083] Practically, the RL agent 302 is responsible for selecting optimal action in the current state according to the knowledgebase. The environment 304 is represented by certain states and a reward function that can output the evaluation of the action selected by the agent at a certain state. After the action is executed, the reward is given by the environment and the agent obtains a numerical reward, and then the current state will move to the next state according to the action; the final objective of the RL agent is to get a most cumulative reward. To realize, the agent must be designed to have the ability to derive the best actions from times of training iterations.

The action-value function under a policy $\pi$ is represented as $Q_\pi(s,a)$, which quantizes the expected return of action $a_t = a$ to be taken in state $s_t = s$ under the policy $\pi$ . In this case, the optimal action-value function $Q^*(s,a)$ is $max_\pi Q_\pi(s_t,a_t)$. when taking best and optimal action in that state can be found out using $a^*(s)$ which is given in (35):

$$a^*(s) = arg\max_a Q^*(s,a) \qquad (35)$$

[0084] Deep Deterministic Policy Gradient (DDPG) is an off-policy method that learns a Q-function and a policy to iterate over actions. It employs the use of off-policy data and the Bellman equation to learn the Q function which is in turn used to derive and learn the policy. Method 1 below corresponds to the optimization of the model 300 shown in FIG.3.

[0085] **Method 1** DDPG method

Randomly initialize critic network $Q(s,a \mid \theta^Q)$ and actor $\mu(s \mid \theta^\mu)$ with weights $\theta^Q$ and $\theta^\mu$.

Initialize target network $Q_t$ and $\mu_t$ with weights $\theta^{Q'}, \theta^Q, \theta^{\mu'}, \theta^\mu$

Initialize replay buffer *R*

**for** episode = 1, M **do**

Initialize a random process $\mathcal{N}$ for action exploration Receive initial observation state $s_1$

**for** t = 1, T **do**

Select action $a_t = \mu(s \mid \theta^\mu) + \mathcal{N}_t$ according to the current policy and exploration noise Execute action *at* and observe reward $r_t$ and observe new state $s_{t+1}$

Store transition $(s_t, a_t, r_t, s_{t+1})$ in *R*

Sample a random minibatch of *N* transitions $(s_i, a_i, r_i, s_{i+1})$ from *R*

Set $y_i = r_i + \gamma Q'\left(s_{i+1}, \mu'\left(s_{i+1} \mid \theta^{\mu'}\right) \mid \theta^{Q'}\right)$

Update critic by minimizing the loss:

$$L = \frac{1}{N}\sum_i \left(y_i - Q(s_i, a_i \mid \theta^Q)\right)^2$$

Update the actor policy using the sampled policy gradient:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s, a \mid \theta^Q)\Big|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s \mid \theta^\mu)\Big|_{s_i}$$

Update the target networks:

$$\theta^{Q'} = \tau\theta^Q + (1-\tau)\theta^{Q'}$$

$$\theta^{\mu'} = \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

**end for**

**end for**

[0086] FIG.4 is a flow chart showing an energy management method for a smart home according to an embodiment of the present invention. The method 400 shown in FIG.4 is carried out by the home energy management system 200 shown in FIG.2.

[0087] In step 402, the interface 214 of the home energy management system 200 receives environmental state indications. The environmental state indications may comprise the electricity price in recent hours, the state of charge of the energy storage system 114 and the electric vehicle 116 and the operating status of the time shiftable load.

[0088] In step 404, the reinforcement learning agent module 222 is executed by the processor 210 of the home energy management system 200 to select a set of control actions based on the environmental state of the smart home.

[0089] In step 406, the control module 228 is executed by the processor 210 of the home energy management system 200 to control the devices of the smart home 100 according to the control actions. Step 406 may comprise the controller 216 of the smart home energy management system 200 generating control signals for the devices of the smart home 100.

[0090] In some embodiments, the method 400 may further comprise the electricity cost forecast module 226 being executed by the processor 210 of the home energy management system 200 to generate predicted energy prices using the electricity cost forecast module 232 stored in the data storage 230.

[0091] The electricity cost forecast module 232 may be trained using historical energy price data. For example, PJM System Energy Price and PJM Hourly Demand (1/3/2019~29/2/2020) data were used to train and test the validation of the proposed neural network.

[0092] FIG.5 shows a comparison of predicted electricity prices calculated by an electricity cost forecast module of an embodiment of the present invention with actual electricity prices.

[0093] In the example shown in FIG.5, the neural network was trained with 6312 data points and tested with 2400 data points, the mean average percentage error shown as following is used as the benchmark of the prediction:

$$MAPE = \frac{1}{n} * \sum_{t=1}^{n} \frac{\left| X_{\text{actual}} - X_{\text{pred}} \right|}{X_{\text{actual}}} * 100\% \qquad (36)$$

[0094] The MAPE of the proposed method is 4.63%, which means the accuracy is above 95%, once the future prices are acquired, they will be regarded as the inputs to the RL methodology described below.

[0095] Parameters of components in the smart home studied in this paper are listed in the table below. These parameters can be manually changed and do not affect the analysis and interpretation of the simulation results.

| Load | Power Rating | Operating Time | α | Efficiency |
|---|---|---|---|---|
| DL1 | 0~3 | 24h | 1 | - |
| NSL2 | 3 | 24h | - | - |
| PSL3 | 0~1 | 7am~3pm | 0.8 | - |
| PSL4 | 0~1 | 24h | 0.8 | - |
| TSL5 | 0.5,0.5,0.7,0.8 | 24h | 2 | - |
| ESS | -2.0~2.0 | 24h | - | 0.9 |
| TL | 0~3 | 24h | - | - |
| TES | -2.0~2.0 | 24h | - | 0.9 |
| mCHP | 0~5 | 24h | - | E0.6&H0.35 |
| EHP | 0~1 | 24h | - | 0.95 |
| EV | 0~10 | 6pm~8am | - | 1 |

[0096] To demonstrate the performance of the proposed HEMS, the detailed simulations based on PJM price of April 6th, 2016 are discussed below. As the case is the simulation in one day, reward of all hour should be treated equally, we set the discounting factor to 1, both actor and critic are 3 hidden layers BPNN, detail parameters are given in the table below.

| Parameter | Value |
|---|---|
| Actor learning rate | 0.001 |
| Critic learning rate | 0.001 |
| L2 regularization factor for the actor | 0 |
| L2 regularization factor for the critic | 0 |
| Learning rate decay (per episode) | 1 |
| Dropout rate for actor | 0 |
| Dropout rate for critic | 0 |
| Number of episodes | 30000 |
| Soft target update rate | 0.01 |
| Number of steps to run the policy (and collect experience) before updating network weights | 8 |

| | |
|---|---|
| Capacity of experience replay memory | 2e7 |
| Size of minibatch from experience replay memory for updates | 128 |
| Pretrain episodes number | 10000 |
| Decay rate (per episode) of the scale of the exploration noise process | 0.9995 |
| μ parameter for the exploration noise process | 0 |
| θ parameter for the exploration noise process | 0.15 |
| σ parameter for the exploration noise process | 0.2 |

[0097] In the training process, the agent can gradually converge to the approximately optimal action choosing.

[0098] FIG.6 shows the rewards of each training iteration in an example embodiment of a home energy management system according to an embodiment of the present invention. At the first 24000 pretraining iterations, the agent randomly chooses the action and the total rewards of each training episode randomly fluctuate at around -5000. After pretraining, the random noise is gradually decreasing in each training episode, with each successive iteration, the rewards increased as the agents discovered the actions that can get higher rewards by learning them through trial and error.

[0099] After the training process, the home energy management system was able to maximise the cumulative rewards.

[0100] FIG.7 is a graph showing electrical balance of each hour for a smart home implementing a home energy management system according to an embodiment of the present invention.

[0101] FIG.8 is a graph showing thermal energy balance of each hour for a smart home implementing a home energy management system according to an embodiment of the present invention.

[0102] On a certain day, the electricity price is fluctuating while the gas price does not change with time. This makes the proposed method have the potential to do an energy-saving decision. Compare with only consume electricity and gas buying from= utilities, the proposed multi-energy smart home is able to utilize the mCHP to consume fuel cell generate more electricity and charge the BESS when the electricity price is high. When the electricity price is low, HEMS will recommend EHP to use electricity to generate thermal energy for home thermal load consumption and thermal storage charging. This bi-directional energy exchange effectively reduces user's daily total energy cost and reduce the electricity consumption in peak hour. By scheduling mCHP, EHP in multi-energy smart home, the user's total energy cost is saved 4.86% in the given day and the monthly average energy cost saving percentage reach 4.69%. When the electricity price is rather high, for example, 18 to 23 pm, the CCHP is more active so that the smart home will consume more natural gas to generate both thermal and electricity. When the electricity price is rather low, like 1 to 7 am, the EHP will be more active to consume electricity to fulfil thermal loads consumption, this will significantly lower the daily total energy cost.

[0103] The table below shows an estimate of the cost saving by the multi-energy structure managed by a home energy management system according to an embodiment of the present invention.

| | 1 day | 5 days | 30 days |
|---|---|---|---|
| Cost saved by CCHP and EHP | 0.09 | 0.47 | 2.83 |
| Percentage | 4.86% | 4.73% | 4.69% |

[0104] FIG.9 is a graph showing an example of energy usage in a smart home without a home energy management system.

[0105] FIG.10 is a graph showing an example of energy usage in a smart home with a home energy management system according to an embodiment of the present invention.

[0106] Comparing FIG.9 and FIG.10, it can be seen that, all the energy demands of controllable and power shiftable loads are scheduled to off-peak slots, i.e., all the controllable loads consume less energy during time slots 18-22pm, and consume more energy during time slots 1-6 and 23-24; the time shiftable load is shifted from peak hour to non-peak hour when the home energy management system is employed. Thus, the overall energy consumption is maintained at a low level during peak slots.

[0107] Example costs for each appliance in the two scenarios is shown in the table below.

| Appliance | Cost without EMS | Cost with EMS | Saving Percentage | 30 days average saving percentage |
|---|---|---|---|---|
| DL1 | 0.447 | 0.425 | 4.3 | 4.6 |
| NSL2 | 0.139 | 0.139 | 0 | 0 |
| PSL3 | 0.083 | 0.022 | 74.1 | 63.5 |
| PSL4 | 0.205 | 0.103 | 49.8 | 43.2 |
| TSL5 | 0.052 | 0.048 | 7.7 | 15.1 |
| Total | 0.927 | 0.738 | 20.4 | 18.8 |

[0108] FIG.11 is a graph showing EV charging scheduling given by a home energy management system according to an embodiment of the present invention. As shown in FIG.11, EV charging is scheduled at low price periods between departure time and arrival time to reduce user's cost. The daily cost of the EV is 0.508.

[0109] FIG.12 is a graph showing state of charge for a battery energy storage system and thermal energy storage system in a smart home using an energy management system according to an embodiment of the present invention.

[0110] FIG.12 shows the action and SOC of BESS and TES in each time slot. Considering the battery long-term degradation cost of BESS, an estimation is given in the table below.

| Cost | Value |
|---|---|
| Battery Capacity | 3 |
| Battery Changing Cost | 50 |
| Battery Life Cycle | 10000 |
| Gain through Charging and Discharging (30 days) | 8.6 |
| Long Term Battery Degradation Cost in One Day | 3.24 |
| Battery Degradation Percentage in One Day | 0.195% |
| Estimate Time to Earn Back Battery Cost (Days) | 174 |

[0111] In 30-day testing, the total battery long-term degradation cost is around 3.24. As the simulation show, the residential user in this case needs around 174 days to get back the investment cost. With the HEMS proposed in this paper, users can evaluate their own electricity consumption behaviour and electricity price policy when investing in BESS, thereby calculating the return on investment. Meanwhile, another two simulations that set BESS or TES capacity to 0 is deployed to evaluate the benefit of installing the energy storage system. For BESS, compare with zero capacity, installing the BESS increase system daily accumulated reward 0.57, which includes reducing daily energy cost by charging and discharging under dynamic price tariff and reducing user's dissatisfactory penalty. For TES, compare with zero capacity, installing the TES increase system daily accumulated reward 0.17 by reducing the user's dissatisfactory penalty.

[0112] As described above, the present disclosure provides a day-ahead home energy management system for a multi-energy smart home based on a deep reinforcement learning method aiming to minimize the total cost function including daily energy cost and user's discomfort level. Electricity prices are predicted by a neural network day ahead. Loads are classified and modelled according to their operating characteristics. mCHP, EHP and EV are scheduled to reduce daily energy cost. Both BESS and TESS are dispatched to optimally perform energy arbitrage for optimal revenues and long-

term battery degradation costs for the battery is considered. Simulation results show that the user's daily electricity price is significantly reduced with the proposed method.

[0113] Whilst the foregoing description has described exemplary embodiments, it will be understood by those skilled in the art that many variations of the embodiments can be made within the scope and spirit of the present invention.

CLAIMS

1. A home energy management system for a smart home, the smart home having an electrical grid connection, a renewable energy source which provides a time varying electrical power output, a battery energy storage system, and a plurality of electrical loads, the home energy management system comprising:

a processor which implements a reinforcement learning agent trained to select control actions for the battery energy storage system and the plurality of electrical loads which minimize an objective function dependent on an energy cost parameter, a comfort metric of the smart home and a degradation metric of the battery energy storage system; and

a controller configured to control the battery energy storage system and the plurality of electrical loads according to the control actions selected by the reinforcement learning agent.

2. The home energy management system according to claim 1, wherein the reinforcement learning agent is a deep deterministic policy gradient agent.

3. The home energy management system according to any preceding claim, wherein the plurality of electrical loads comprises a power shiftable load.

4. The home energy management system according to any preceding claim, wherein the plurality of electrical loads comprises a time shiftable load.

5. The home energy management system according to any preceding claim, wherein the plurality of electrical loads comprises an electrical vehicle charging load.

6. The home energy management system according to claim 5, wherein the electrical vehicle charging load has a specified charging period in which an electrical vehicle must be charged.

7. The home energy management system according to any preceding claim, wherein the processor further implements an energy cost forecast module configured to predict energy cost.

8. The home energy management system according to claim 7, wherein the energy cost forecast module comprises an artificial neural network trained to predict energy cost.

9. The home energy management system according to any preceding claim, wherein the smart home has a gas input connection and further comprises a gas boiler configured to generate thermal energy from gas, and wherein the reinforcement learning agent is further trained to select control actions for the gas boiler to minimize the objective function.

10. An energy management method for a smart home, the smart home having an electrical grid connection, a renewable energy source which provides a time varying electrical power output, a battery energy storage system, and a plurality of electrical loads, the energy management method comprising:

receiving environmental state indications indicating an environmental state of the smart home;

selecting, by a reinforcement learning agent, control actions for the battery energy storage system and the plurality of electrical loads which minimize an objective function dependent on an energy cost parameter, a comfort metric of the smart home and a degradation metric of the battery energy storage system; and

controlling the battery energy storage system and the plurality of electrical loads according to the control actions selected by the reinforcement learning agent.

11. The energy management method according to claim 10, wherein the reinforcement learning agent is a deep deterministic policy gradient agent.

12. The energy management method according to any one of claims 10 to 11, wherein the plurality of electrical loads comprises a power shiftable load.

13. The energy management method according to any one of claims 10 to 12, wherein the plurality of electrical loads comprises a time shiftable load.

14. The energy management method according to any one of claims 10 to 13, wherein the plurality of electrical loads comprises an electrical vehicle charging load.

15. The energy management method according to claim 14, wherein the electrical vehicle charging load has a specified charging period in which an electrical vehicle must be charged.

16. The energy management method according to any one of claims 10 to 15, further comprising predicting future energy costs.

17. The energy management method according to claim 16, wherein predicting future energy costs comprises using an artificial neural network trained to predict energy cost.

18. The energy management method according to any one of claims 10 to 17, wherein the smart home has a gas input connection and further comprises a gas boiler configured to generate thermal energy from gas, and wherein the energy management method further comprises selecting control actions for the gas boiler to minimize the objective function.

19. A computer readable medium storing processor executable instructions which when executed on a processor cause the processor to carry out a method according to any one of claims 10 to 18.