Home Energy Recommendation System (HERS): A Deep Reinforcement Learning Method Based on Residents' Feedback and Activity

Salman Sadiq Shuvo, *Graduate Student Member, IEEE*, and Yasin Yilmaz, *Senior Member, IEEE*

***Abstract*—Smart home appliances can take command and act intelligently, making them suitable for implementing optimization techniques. Artificial intelligence (AI) based control of these smart devices enables demand-side management (DSM) of electricity consumption. By integrating human feedback and activity in the decision process, this work proposes a deep Reinforcement Learning (RL) method for managing smart devices to optimize electricity cost and comfort residents. Our contributions are twofold. Firstly, we incorporate human feedback in the objective function of our DSM technique that we name Home Energy Recommendation System (HERS). Secondly, we include human activity data in the RL state definition to enhance the energy optimization performance. We perform comprehensive experimental analyses to compare the proposed deep RL approach with existing approaches that lack the aforementioned critical decision-making features. The proposed model is robust to varying resident activities and preferences and applicable to a broad spectrum of homes with different resident profiles.**

***Index Terms*—Home energy management, demand side management, customer comfort, residents' activity label, deep reinforcement learning, artificial intelligence.**

Nomenclature

Devices

AC	Air Conditioning
DW	Dish Washer
WD	Washer and Dryer
EV	Electric Vehicle.

Parameters at Time Step t

X_t^i	Activity label of the ith resident.
H_t^i	Previous activity labels of the ith resident.
ω_t^i	Duration of current activity in hours of the ith resident.
ρ_t	Real-time electricity price in \$/kWh.
f_t^d	$\in \{0, 1\}$, Human feedback for device d.
P_t^d	Power consumption in kW for device d.
β_t^d	$\in [0, 1]$, Battery charge level of device d.

Other Symbols

D	Total number of smart devices.
d	Device index.
δ_d	Discomfort cost coefficient for device d.
ψ_d	Scheduling time steps for device d (Type-2).
$\mathcal{E}_d$	Charging power levels for device d (Type-3).
κ	Duration of time step in hours.

I. Introduction

A. Home Energy Management (HEM) for Smart Homes

Smart home systems can enhance human comfort and optimize electricity usage in an automated setup. While many devices have included sensor-based control for a long time, such as microwave ovens, air conditioning, etc., with the Internet of Things (IoT) revolution [1], many other smart appliances are entering our homes. Most of the devices will soon have such intelligence that will unlock the true potential of the smart home concept. Specifically, recent smart Home Energy Management (HEM) technologies can leverage state-of-the-art artificial intelligence (AI) techniques. As a result, residents can enjoy all the comfort smart devices offer according to their preferences in an automated way. In addition to personalized comfort, the HEM system can significantly reduce the electricity cost and flatten the demand curve by scheduling some devices to run during off-peak hours.

B. Demand-Side Management (DSM) Techniques for HEM

Utility companies employ Demand Response (DR) based techniques to encourage customers to shift their load to off-peak hours [2]. It serves two purposes: avoiding electricity purchases from expensive peaking power plants and keeping the system's maximum demand at check to avoid capacity expansion costs. They provide time-based pricing schemes for the customers, known as Time of Use (TOU) [3], such as real-time pricing, critical peak pricing, etc. Numerous researches have proposed appliance scheduling techniques for HEM systems [4] to capitalize TOU tariffs. Such Demand Side Management (DSM) techniques aim to modify the consumer's energy activities, e.g., shifting customers' electricity usage towards off-peak hours [5]. For instance, a hierarchical HEM system within a home microgrid is proposed in [6] that integrates photovoltaic (PV) energy into day-ahead load scheduling and aims to reduce the monthly peak demand and

Manuscript received August 29, 2021; revised December 6, 2021 and January 24, 2022; accepted March 7, 2022. Date of publication March 11, 2022; date of current version June 21, 2022. Paper no. TSG-01385-2021. *(Corresponding author: Salman Sadiq Shuvo.)*

The authors are with the Department of Electrical Engineering, University of South Florida, Tampa, FL 33620 USA (e-mail: salmansadiq@usf.edu; yasiny@usf.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2022.3158814.

Digital Object Identifier 10.1109/TSG.2022.3158814

peak demand charges.[1] A state-space approximate dynamic programming (SS-ADP) approach is proposed in [7] to provide a fast real-time control strategy under uncertainty using the Bellman optimality condition. The work in [8] includes consumer input in their proposed EV charge scheduling technique. The uncertainties in electricity usage of smart building HEM as a nonlinear optimization problem is addressed in [9]. A microgrid where the users minimize cost by trading energy between each other before buying from the grid is presented in [10], where PV energy, home battery, and EV battery serve as intermittent sources.

The majority of the DSM techniques for HEM are based on a rule-based schedule for device usage, undermining consumers' comfort. Rule-based scheduling often suffers from the randomness inherent in human preference, weather, and other interventions, especially in realistic scenarios with multiple residents and multiple appliances. To this end, the works in [11]–[13] aim to dissolve the rigid scheduling of devices by including distributed energy generation and distributed energy storage devices in their HEM system. To realize the far-reaching potential of smart home technology, researchers have opted from rule based approaches to recent data-driven machine learning techniques for DSM.

C. Reinforcement Learning (RL) Based DSM Techniques

Electricity consumption patterns are evolving with the fast-improving smart device technologies, which requires adaptability in HEM for scheduling devices. Reinforcement learning (RL) techniques are typically preferred for their data-driven online decision-making capability. Recent advances in neural network-based deep RL algorithms lead to widespread applications, including gaming [14], finance [15], energy systems [16], transportation [17], communications [18], environmental systems [19], and healthcare systems [20]. An extensive review of RL for DSM in [21], showcases the suitability of RL for DSM techniques. Berlink and Costa [22] were among the first to investigate RL-based DSM techniques for a smart home. The work [23] utilizes the inherent adaptability in deep RL algorithms by maintaining thermal comfort and optimal air quality while minimizing electricity usage. A large-scale HEM is proposed in [10] using a multi-agent deep RL framework.

D. Human Feedback and Activity for HEM

The authors, in their review of RL for demand response [21], emphasize the importance of incorporating human feedback in RL-based DSM techniques. Pilloni *et al.* [24] proposed a smart HEM system in terms of the quality of experience, which depends on the information of consumers' discontent for changing home devices' operations. To replicate human feedback, they surveyed 427 people to generate residents' annoyance profiles for delayed scheduling of different appliances. Then, they incorporate a cost apart from the electricity price based on the annoyance levels from these profiles. In their following research [25], they used sensor-based activity recognition to predict future activities for appliance scheduling. The authors in [26] define human dissatisfaction by the difference between the maximum power rating and the delivered power rating of a device, an oversimplified way of representing human feedback for their RL-based HEM system. Khan *et al.* [27] calculated dissatisfaction if HEM turns off a device using an equation with different priority factors for different devices. Several other works, e.g., [28], [29], follow a similar approach to estimate discomfort cost rather than using actual feedback from residents. All these techniques lack adaptability to consumer preference, i.e., they may work well for certain types of users, but they are not general enough to ensure user convenience. Park *et al.* [30] provide theory and implementation for adaptive and occupant-centered lighting optimization in an office setup. They interpret switching on and off the lights by office employees as human feedback. This work has successfully incorporated human feedback for their RL algorithm; however, their scope is limited to lighting. Hence, the necessity for a human feedback-based HEM system still remains open.

The work in [31] proposes a deep sequential learning-based human activity recognition in smart homes. The benefits of labeled activity to analyze and assess the smart home residents' physical and psychological health has been reviewed in [32]. Chen *et al.* [33] analyzed behavior patterns to predict energy consumption profile. Since the smart home concept has the inherent capability of activity labeling, including the activity data as a feature for the DSM technique can greatly facilitate the RL agent's learning capacity. The work [34] reviews sensor-based activity recognition techniques to implement in a smart home setup. Given the technology, our work includes human activity labels in the RL state definition for the first time to the best of our knowledge.

Although the smart home concept is originally introduced for the residents' benefit, their comfort is often ignored in many existing methods. In this work, we propose a deep RL method that takes the residents' feedback as a reward factor, apart from electricity prices and device status. We consider resident activities as part of the system state to better understand human comfort and feedback. Our work incorporates residents' feedback every time they override the HEM system's commands, a practical and novel way of extending the success of recommender systems (e.g., movie, book, shopping, video) to HEM. Recommender systems learn from customer usage patterns to recommend items/services [35]. A similar approach can be integrated into a HEM system by accommodating human input in a meaningful way.

E. Contributions

Our contributions lie in addressing two challenges in RL for HEM. Specifically,

- We propose a novel home energy recommender system (HERS) based on a Markov decision process (MDP) formulation and a deep RL solution to jointly minimize the electricity consumption cost and discomfort to the residents;

[1] Not every utility charges for peak demand.

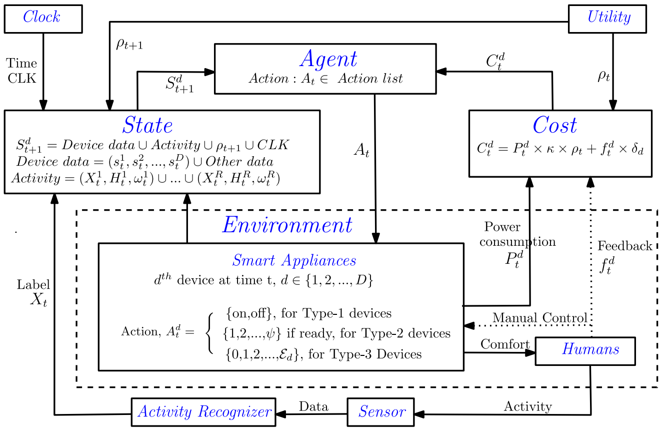

Clock
Utility
Time CLK
ρ_{t+1}
Agent
Action : $A_t \in$ Action list
S^d_{t+1}
C^d_t
ρ_t
State
$S^d_{t+1} = Device\ data \cup Activity \cup \rho_{t+1} \cup CLK$
$Device\ data = (s^1_t, s^2_t, ..., s^D_t) \cup Other\ data$
$Activity = (X^1_t, H^1_t, \omega^1_t) \cup ... \cup (X^R_t, H^R_t, \omega^R_t)$
A_t
Cost
$C^d_t = P^d_t \times \kappa \times \rho_t + f^d_t \times \delta_d$
Environment
Power consumption P^d_t
Feedback f^d_t
Label X_t
Smart Appliances
d^{th} device at time t, $d \in \{1, 2, ..., D\}$
Action, $A^d_t = \begin{cases} \{\text{on,off}\}, \text{ for Type-1 devices} \\ \{1,2,...,\psi\} \text{ if ready, for Type-2 devices} \\ \{0,1,2,...,\mathcal{E}_d\}, \text{ for Type-3 Devices} \end{cases}$
Manual Control
Comfort
Humans
Activity Recognizer
Data
Sensor
Activity

Fig. 1. Proposed MDP model.

TABLE I
DEVICE TYPES

Devices	Priority (Type-1)	Deferrable (Type-2)	Flexible (Type-3)
Device Properties	Not deferrable, Rigid power consumption	Deferrable, Rigid power consumption	Deferrable, Flexible power consumption
Operational Objective	Minimize idle usage by turning on/off intermittently	Operate during low ρ_t	Charge highly at low ρ_t
Device Ready Status	Turned on by resident/sensor	Turned on by resident	Connected to charger
Action Selection	Every time step during active period	Once every activation	Every time step during active period
Actions	$A^d_t \in$\{on,off\}	$A^d_t \in \{0, 1, 2, ...,\psi_d \}$	$A^d_t \in \{0, 1, 2, \ldots, \mathcal{E}_d\}$
Device Examples	Regular lights, TV, AC	Sprinkler, DW, WD	EV, cell phone, laptop chargers

- HERS incorporates *direct human feedback for discomfort* in the objective function through residents' manual overrides to the recommended device operations to learn residents' preferences; and
- HERS uses *resident activities* in the state definition to learn device usage patterns.

We evaluate the performance of the proposed HERS method by comparing with a manually controlled, two rule-based [13], [24], and an RL-based approach [26].

The remainder of the paper is organized as follows. The MDP model is formulated in Section II, and the deep RL algorithm for the optimal policy is given in Section III. The experimental setup is presented in Section IV. Results are discussed in Section V. Finally, after the key features of the paper and the future research scope are discussed in Section VI, the paper is concluded in Section VII.

II. MODEL DEVELOPMENT

We propose an MDP framework shown in Fig. 1, where the smart home device manager is the MDP agent, called HERS.

A. Environment

The residents, activity recognizers, and devices form the MDP environment. The homes can be of different sizes, with multiple residents living in them. We assume access to the utility company's real-time pricing scheme, ρ_t (\$/kWh at time t), and activity recognition through multiple sensors placed throughout the home. Affordable and reliable activity recognition from sensor data has been studied by several works [31]–[33], which is out of the scope of this paper. We assume the presence of an activity recognition set up, which provides the activity label $X^1_t, X^2_t, \ldots, X^R_t$ for all R residents at home.

HERS employs different methods to operate each of the $d \in \{1, 2, \ldots, D\}$ smart devices that we divide into three categories, as shown in Table I. When switched on by a human or sensor, the device goes into the active status and will be considered for decision-making only during active status.

1) Priority Devices (Type-1): These devices provide essential comfort to the residents, and they are not available for deferring. HERS can keep the active devices off intermittently without compromising the devices' functionality. Regular lights, TV, CCTV camera, alarm system, and air conditioner (AC) are examples of this type of appliance. Choosing the relevant data for the MDP state is a challenge for this task. For instance, if the resident is browsing the Internet while the TV is on, turning it off may create discomfort. However, if the resident goes to sleep, keeping the TV on, turning it off may reduce electricity costs without compromising comfort. AC is the heaviest load for this device type, hence we focus on it in our experiments.

2) Deferrable Devices (Type-2): These devices can be scheduled later to off-peak hours, reducing electricity cost and maintaining the peak demand lower than the threshold (if any). Dish Washer (DW) and Washer & Dryer (WD) fall in this category. These devices typically can evade human discomfort if it completes the task before the subsequent activation by the residents. So, the dynamic electricity price ρ_t and activation time are critical features for scheduling the deferrable devices.

3) Flexible Devices (Type-3): These devices are flexible in terms of time scheduling and power level. EV, cell phone, and laptop chargers are examples of these types of devices. These devices can consume different power levels $\{0, 1, 2, \ldots, \mathcal{E}_d\}$, which changes their battery charge level β^d_t. Residents' activity patterns and β^d_t are important features in HERS for these devices.

B. Action, A^d_t

Our MDP model in Fig. 1 starts with the agent selecting actions $A_t = (A^1_t \cup A^2_t \cup \cdots \cup A^{D_{act}}_t)$ about setting the operation mode for each of the smart devices in active status D_{act} ($\leq D = m + n + o$). So, the total number of possible actions are

$$\underbrace{2^m}_{m\ Type-1\ Devices} \times \underbrace{(\psi_1 + 1) \times (\psi_2 + 1) \times \cdots \times (\psi_n + 1)}_{n\ Type-2\ Devices} \times \underbrace{(\mathcal{E}_1 + 1) \times (\mathcal{E}_2 + 1) \times \cdots \times (\mathcal{E}_o + 1)}_{o\ Type-3\ Devices}, \tag{1}$$

where, m is the total number of Type-1 devices. $\psi_1, \psi_2, \ldots, \psi_n$ are scheduling time ranges for the n type-2 devices, and $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_o$ are charging power levels for the o type-3 devices. Fig. 2 shows the action flowchart for each type of devices at each time t.

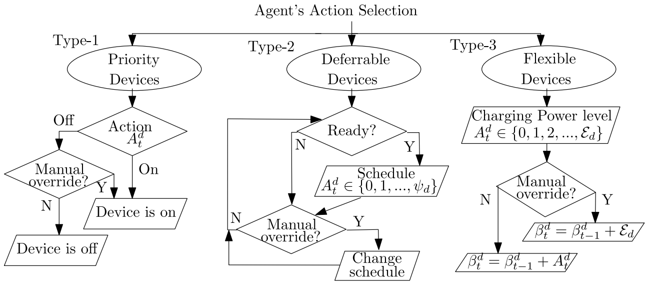


Fig. 2. Action flow chart for active devices at each time t.

TABLE II
STATE INPUT

Input	AC	DW	WD	EV
Activity	✓	✓	✓	✓
Clock time	✓	✓	✓	✓
Electricity price, ρ_t	✓	✓	✓	✓
Device status	✓	✓	✓	✓
Device activation duration	✓	✓	✓	✓
Battery charge level	×	×	×	✓
Travel upcoming	×	×	×	✓

For the type-1 devices, there are two actions possible (on/off) for the device. For A_t^d = off, the agent changes its action if the residents' perform manual override. HERS schedules a Type-2 device when it is ready for a new run. No further decision is made until the current operation is finished, either scheduled or manually overridden. The device becomes ready again when the resident activates it for a new run. For Type-3 devices, HERS decides on a charge level A_t^d for each time t. If there is any manual override, then the charge level is set to full capacity $\mathcal{E}_d$ to finish charging as soon as possible. Every manual override causes the discomfort cost through feedback $f_t^d = 1$ to the RL agent for the corresponding device.

The actual number of possible actions will typically be smaller than Eq. 1 during a time step due to inactive devices. For example, when the residents are not at home, the AC will remain off and will not be considered for the agent's action. Similarly, the idle status of many devices can be determined to limit the number of actions. Furthermore, a deferrable device only remains active for one time step when HERS schedules its operation.

C. State, S_t^d

The MDP agent takes action based on the environment state. Appropriate design of the state is fundamental to the success of the MDP model. As the devices provide comfort to the residents, we hypothesize their activity data to be critical to define the states. An activity recognition system uses various home sensor data to label the residents' activity X_t. Apart from the activity, the real-time electricity price ρ_t and clock time of the day (CLK) are other essential features that we include in the state definition, as shown in Table II. The state at time t is defined as:

$$S_t^d = \text{Device data} \cup \text{Activity} \cup \rho_t \cup CLK,$$

where Activity $= (X_t^1, H_t^1, \omega_t^1) \cup \cdots \cup (X_t^R, H_t^R, \omega_t^R)$ includes the current activity X_t^i, previous activities H_t^i, and duration of the current activity ω_t^i for all R residents. Device data includes information like how long ago the device was activated, the number of dirty dishes or clothes for the Dishwasher and Washer Dryer, the charge level of the type-3 devices, that can be included in the state definition, as shown in Table II. In practice, activity labels can be generated from activity recognition sensors as discussed in [31].

D. Cost, C_t^d

The MDP agent tries to maximize a reward or minimize a cost by taking optimal actions for a given state. For instance, the RL-based Youtube video recommendation systems are rewarded when the user opens a recommended video [36]. Similarly, HERS receives cost (negative reward) whenever a resident is not happy with the selected action and changes the mode of a device. This human feedback $f_t^d = 1$ is interpreted as discomfort and converted to a cost to the MDP agent through separate cost coefficients δ_d for each device d for each manual override. The devices' operations are meant for human comfort, so HERS' objective is to minimize discomfort.

The total cost for the MDP agent is the sum of energy cost and human discomfort cost. The utility informs the agent of the electricity price for the current time step ρ_t, and future time step ρ_{t+1}. The energy usage at time t is obtained from the smart device's power consumption P_t^d and used to calculate the total cost for each active device for time step t as

$$C_t^d = P_t^d \times \kappa \times \rho_t + f_t^d \times \delta_d, \tag{2}$$

where κ is the unit step time in hours. Cost coefficient δ_d for each device is a critical modeling parameter that converts discomfort into monetary value. f_t^d represents the discomfort feedback of the residents, where 0 and 1 respectively indicates no override or override. The goal of the MDP agent is to minimize the following discounted cumulative cost for each device in T time steps:

$$C_T^d = \sum_{t=0}^{T} \lambda^t C_t^d, \tag{3}$$

where $\lambda \in [0, 1]$ is the discount factor for future decisions.

E. Next State, S_{t+1}^d

At the end of a time step, the device state S_t^d changes according to the action A_t; however, human activity data, electricity price data, etc., change stochastically. These features define the next state S_{t+1}^d, and the dynamic system moves to the next time step for the agent to act. These transitions satisfy the Markovian property of the MDP framework.

III. SOLUTION APPROACH

HERS employs one separate MDP agent for each of the D devices to minimize the discounted total costs C_T^d in Eq. (3). To achieve the optimal policy $\arg\min_{\{A_t^d\}} C_T^d$, we need to solve the following Bellman equation. We drop the device index from here on for brevity. The agent's value function at time step t is

$$V(S_t) = \min_{A_t} \big\{ \mathsf{E}\big[C_t + \gamma V(S_{t+1})\big]\big\}.$$

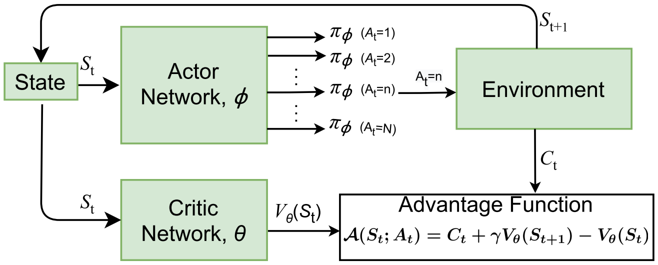


Fig. 3. Advantage Actor-Critic (A2C) Network.

The above equation presents a solution dilemma in prioritizing between the immediate cost C_t and future expected cost $\gamma V(S_{t+1})$. Since the agent's action changes the next state of the devices, the future discounted cost through the value function of the next state $V(S_{t+1})$ depends on the action of the agent. Since in high-dimensional problems like the considered one here, it is not feasible to compute the expected future cost explicitly and find the value function for each possible state, deep neural networks are typically used in the modern practice of RL (known as deep RL) to learn the optimal policy of actions either directly (policy-based methods) or through the value function (value-based methods).

The Advantage Actor-Critic (A2C) algorithm, which is a hybrid (both value-based and policy-based) adaptation of policy gradient-based algorithm REINFORCE [37], is a popular choice for continuous state space, e.g., electricity price and battery charge level in our setup. We also considered using Deep Q Network (DQN), another popular deep RL algorithm, but A2C performed better in the proposed state space, as expected. A2C uses the advantage functions for policy update, which reduces the REINFORCE algorithm's variance as shown in Fig. 3.

The actor network, also known as the policy network, outputs probability for each action value $\pi_\phi(A_t)$ through a softmax function. Then the agent samples an action A_t based on the policy π_ϕ and the environment moves to the next state S_{t+1} and provides the immediate cost C_t. The actor network aims to find the gradient of expected return $J(\pi_\phi)$ of the policy π_ϕ with respect to the weights ϕ of the neural network through the following equation:

$$\nabla_\phi J(\pi_\phi) = \mathsf{E}_{\pi_\phi}\left[\nabla_\phi \log\left(\pi_\phi(A_t|S_t)\right)\mathcal{A}(S_t;A_t)\right], \tag{4}$$

where the advantage function $\mathcal{A}$ is given by

$$\mathcal{A}(S_t;A_t) = C_t + \gamma V_\theta(S_{t+1}) - V_\theta(S_t). \tag{5}$$

The critic-network learns the value function $V_\theta(S_t)$ for each state. It uses the advantage function $\mathcal{A}$ as the critic loss to update its network parameters θ through back propagation. A pseudo code for the proposed A2C algorithm is given in Algorithm 1.

Algorithm 1 A2C Algorithm for Each Device in HERS

```
Input: discount factor λ, discomfort cost coefficient $\{\delta_d\}$
Initialize: Actor network with random weights $\phi$ and critic network with random weights $\theta$
for episode = $1, 2, \ldots, E$ do
    for t = $1, 2, \ldots, T$ do
        Collect activity data from Activity Recognizer, real-time electricity price $\rho_t$ from Utility.
        Select action $A_t^d$ using Actor Network (Fig. 3).
        Execute action $A_t^d$ and observe human discomfort feedback $f_t^d$.
        Calculate cost $C_t^d$ using Eq. (2).
        Store transitions $(S_t^d, A_t^d, C_t^d, S_{t+1}^d)$.
    end for
    Update actor network $\phi$ via Eq. (4).
    Update critic network $\theta$ through back propagation.
end for
```

TABLE III
ARAS Activity Dataset [39]

	House A	House B
Size	$538ft^2$	$969ft^2$
Layout	One bedroom, one living room one kitchen, one bathroom	Two bedrooms, one living room one kitchen, one bathroom
Residents	2 males at their twenties	Married couple at their thirties
Duration	30 days	30 days
Published in	2013	2013
Labelled Activities: (1) Going out, (2-4) Cooking (breakfast, lunch, dinner), (5-7) Having breakfast, lunch, dinner, (8) Washing dishes, (9) Having snack, (10) Sleeping, (11) Watching TV, (12) Studying, (13) Bath, (14) Toileting, (15) Napping, (16) Using Internet, (17) Reading book, (18) Laundry, (19) Shaving, (20) Brushing teeth, (21) Phone conversation, (22) Listening music, (23) Cleaning, (24) Conversation, (25) Having guest, (26) Changing clothes, (27) Other		

IV. Experimental Setup

The ideal experimental setup would be implementing the HERS algorithm in an existing smart home. However, a fully equipped smart home capable of taking human feedback is yet to be available. We will hypothetically generate human feedback and interactions with the devices based on the residents' activity data. HERS select different features for operating different devices, as shown in Table II. We include clock time in minutes and real-time electricity price ρ_t as the common states for all the devices. The New York Independent System Operator (NYISO) provides real-time electricity prices; we use Long Island, NY prices for March 13 and 19, 2021 as the electricity price respectively for weekends and weekdays in our simulation [38]. We find that $\kappa = 0.25$ hour (15 minutes) is suitable for the experimental setup.

A. Activity Label

For residents' indoor activity data, we use the ARAS dataset [39]. The attributes of the dataset for the two homes are shown in Table III. The dataset contains 27 types of activities labeled by sensors and validated by the residents. This dataset is comparatively newer and has more activity types than other datasets in the literature.We choose House B for the experiments. HERS takes the current activity label, duration, and the last activity label for each resident (6 inputs in total for the two residents in the house). Apart from providing the dataset, [39] also gives a guideline about the sensors required for activity recognition. To collect the activity data, they used a total of 20 binary sensors of 7 types: (1) force sensor, (2) photocell,

(3) contact sensors, (4) proximity sensors, (5) sonar distance sensors, (6) temperature sensors, and (7) infrared sensors.

B. Devices

HERS can provide optimal control for all the smart devices in a home. However, we limit our case study to high power loads that renders significant energy cost. Specifically, we choose the following four devices for our experiments.

1) Type-1, Central AC: We estimate a 12000 BTU (3.5 kWh) AC capacity for the 90 m^2 (968.75 ft^2) area of the home, located in a mild temperature zone. In reality, the average AC load is typically half of the capacity [40], so we model the AC load with the following normal distribution:

$$P_{AC} \sim \mathcal{N}(\mu = 1.8 \text{ kW}, \sigma = 0.5 \text{ kW}).$$

The AC will be in the idle status ($s_t^{AC} = 0$) if none of the residents are at home or active ($s_t^{AC} = 1$) otherwise. The agent may keep the AC off intermittently under active status; however, the resident will manually turn the AC on if it causes discomfort. We generate this feedback f_t^{AC} if AC goes off within T_{AC} minutes of being turned on. In that case, the residents turn on the AC manually, which penalizes the HERS agent by \$ δ_{AC}. We model T_{AC} with a uniform distribution between 45 to 90 minutes and intermittent off duration as 15 minutes. We include the on duration as a state for AC.

2) Type-2, DW and WD: The activity pattern of house B indicates the lack or no usage of a dishwasher (DW). We generate the dishwashing events to be activated, i.e., $s_t^{DW} = 1$, T_{DW} minutes after any resident finishes dinner. We model the delay time T_{DW} with the Poisson distribution with 60 minutes mean value. There will be no dishwashing events for the days when none of the residents have cooked, as there will not be a significant load for the dishwasher. The analysis in [41] estimates 152-minute automatic dishwashing for a comparable load to the considered household. Hence, we model the dishwashing event as a 2.5-hour continuous operation with 1.1 kW power. The Bosch 500 series smart dishwashers are among the most popular models of the year 2020 and serve as the DW model in our experiments [42]. The agent needs to complete dishwashing before the subsequent switching by the residents; otherwise, it receives the discomfort cost δ_{DW}, and the DW is turned on manually to clean the previous dishes.

House B has a regular heavy load washer & dryer (WD), so following its laundry schedule would not be practical. The future smart homes will utilize the high-tech WD combos like the LG WM3900HBA, a single compartment light-duty device that takes around 1 hour for washing and 1.5 hours for drying for an average cloth load. We estimate that the residents produce this cloth load every three baths, hence fill and switch the WD in active mode on average 30 minutes (Poisson mean) after their second or third bath (with equal probability) from the previous laundry. Then the RL agent has to turn the WD on for a 1-hour continuous washing cycle, followed by 1.5-hour drying cycles with 1.2 kW power to complete the laundry. If the agent does not complete the process before the next switching by the residents, the resident provides negative feedback δ_{WD} and turns on the WD immediately to clean the previous cloths.

TABLE IV
EV USAGE DATA GENERATION

	Weekdays			Weekends	
Duration, t_a (hrs)	< 8	8-16	> 16	< 10	> 10
Purpose	Leisure	Office	Travel	Leisure	Travel
Miles driven, M	$f(t_a)$	40+$f(t_a$-10)	n/a	$f(t_a)$	n/a
Minimum Battery Before Trip	40%	40%	70%	40%	70%
Battery Status After Trip	$\beta - \frac{M}{220}$	$\beta - \frac{M}{220}$	20%	$\beta - \frac{M}{220}$	20%

3) Type-3, EV Charging: The residents' activity pattern shows that they mostly go out of the home together. Considering an EV in the house, we assume that the second resident drives it. The EV driver's work pattern seems to consist of long hours with some off days throughout the week. We set his one-way drive to work as 20 miles; 69th percentile driving distance from the data collected by The American Time Use Survey (ATUS) [43], which includes over 13,000 respondents. The activity data provides us with the duration the resident is away from home. Based on the duration, we label such away time as leisure, office time, and travel as in Table IV. We assume the EV is always connected to the charger when the resident is at home.

For weekdays, if the resident stays away for less than 8 hours, it is labeled as a leisure activity, which includes going shopping, visiting friends, short trips, theater, etc. Residents spend more time in leisure activities during the weekend, extending the leisure activity labeling time to 10 hours for the weekend. Driving distance in miles during leisure trip for t_a time duration is approximated as;

$$M = f(t_a) = t_{\text{driving}} \times v_{\text{avg}} = \alpha \times t_a \times v_{\text{avg}}.$$

where, $\alpha = t_{\text{driving}}/t_a$ is the ratio of time spent for driving and the total time spent away. We model it with a normal distribution

$$\alpha \sim \mathcal{N}(\mu = 0.33, \sigma = 0.1).$$

Average speed v_{avg} is taken as 30 mph. The instances in which the resident spends 8-16 hours out of home is labeled as office and leisure activity during weekdays. Round trip to the office is taken as 40 miles, additional time after 10 hours is considered a leisure activity, and driving distance is calculated as $M = 40 + f(t_a - 10)$.

2021 Tesla Model 3 Standard Range is one of the most popular latest EV models with a 450 hp (336 kW) engine 50 kWh battery. The level–2 charging of 7.68 kW (240 V 32 A) capacity would require 6.5 hours to charge the completely depleted EV battery fully. Battery status after a trip is the initial battery status when going out of home β minus $\frac{M}{220}$ as the Tesla 3 model has a standard driving range of 220 miles. The resident does not use the EV if β is less than 40% before starting a trip. The resident takes some other transportation mode and assigns a discomfort cost δ_{EV1} to the RL agent. If the resident stays more than 16 and 10 hours out of home, respectively, on weekdays and weekends, we label this activity as travel that may require outside charging. We do not calculate driving distance for traveling; however, we set battery status after the travel to be 5-20%, as home charging is the cheapest and the resident would try outside (paid) charging as little as necessary to reach home. The resident requires a higher initial charge for traveling. We set a higher discomfort cost δ_{EV2} if $\beta < 70\%$ before travel. As the initial charge level is higher for travel, we include the next trip type as an input state for the EV.

C. Discomfort Cost, δ_d

Discomfort costs δ_d for each device are critical parameters in the HERS setup. So, we model it as user-defined numbers that the residents can set initially and update while the HERS is at service. The discomfort costs also represent the comfort and device priority mindset of the residents as low discomfort cost will emphasize electricity cost, and high discomfort cost will prioritize human feedback. In case of an update to the discomfort costs δ_d, thanks to its adaptive nature, the RL agent will update the policy in an online fashion. For the experimental purpose, we performed a survey among twenty participants with different backgrounds (e.g., student, homemaker, engineer, etc.) to set the discomfort cost for each of the four devices. Survey results suggest EV charging failure creates the maximum discomfort. Other discomfort costs in decreasing order are for WD, DW, and AC. We select discomfort cost coefficients as $\delta_{AC} = 20$, $\delta_{DW} = 40$, $\delta_{WD} = 50$, $\delta_{EV1} = 100$, $\delta_{EV2} = 300$ in USD.

V. Results

A. Benchmark Policies

1) Manually Controlled Policy: In this policy, the residents operate the devices themselves, so a device turns on immediately upon its activation without any scheduling consideration. We assume the residents turn off the AC when both of them are out of home and turn it on upon returning. This policy ignores the benefit of smart scheduling, and we will refer to it as the baseline policy to evaluate the other policies' success.

2) Rule Based HEM in [13]: Shirazi and Jadid [13] present a home energy management with DERs and appliance scheduling (HEMDAS). The energy management problem in a house is modeled as a mixed-integer nonlinear programming (MINLP) that includes constrained optimization for managing DERs and appliance usage. More precisely, the devices are scheduled based on real-time pricing of electricity during a time window. They define separate earliest starting times (EST) and latest finish times (LFT) for DW, WD, and EV to ensure user convenience. Each device is scheduled based on the real-time electricity price during its operating time window. The AC maintains the desired temperature decided by the customer, which our smart home agent ensures by keeping the AC on for 90 minutes before every 15-minute interruption.

3) Rule-Based HEM in [24]: Pilloni *et al.* [24] survey 427 people about their degree of annoyance if a device performs under-capacity or is scheduled for later periods. The survey responses are used for generating different types of resident profiles. During training, the smart home residents' usage pattern is matched to one of those profiles. Once the resident's appliance usage profile is assigned, the algorithm minimizes the cost for each device,

$$C_t^d = \frac{P_t^d \times \kappa \times \rho_t}{\sigma(\Delta X)}$$

where the numerator represents the electricity cost and $\sigma(\Delta X) \in (0, 1]$ is the relative satisfaction level of the home residents for the device. This rule-based method accommodates user preference and provides a good analogy to our discomfort feedback-based RL approach. The resident feedback pattern in our setup for the AC, DW, and WD matches most of the resident profiles in the survey. Since [24] does not provide an EV charging profile, we assume that this policy schedules EV only if its battery is more than 50% charged, otherwise charges at full capacity.

TABLE V
Monthly Cost ($) Comparison for Different Policies

Device / Policy	AC	DW	WD	EV		Total	
				S1	S2	S1	S2
Manual Control	122.9	5.06	3.48	62.01	62.16	193.3	193.4
Rule-based in [13]	104.6	4.39	2.35	68.32	70.55	179.6	181.8
Rule-based in [24]	104.6	3.23	2.35	56.23	60.97	166.4	171.1
RL-based in [26]	106.7	3.35	2.41	59.76	60.37	172.2	172.8
Proposed HERS	92.0	3.19	2.3	51.6	55.88	149.1	153.4

4) RL-Based HEM in [26]: In [26], Xu *et al.* utilized hour-ahead electricity price as a state to minimize electricity cost. We tailor their approach to fit this comparative analysis with the following modifications: (i) Agent makes decisions every 15 minutes instead of hourly decisions. (ii) There is no PV generation in our setup, so the MDP state consists of electricity price of the next 24 hours, with 4.67% prediction error following the case-1 (best prediction) in that paper. (iii) We consider the AC as a priority device that maintains the user set the temperature on its own. Hence, the possible actions for the AC remain turn on or off instead of different power ratings, (iv) We include EV battery depletion, which is overlooked in [26].

B. Scenarios

1) Scenario 1 (Unlimited Peak Demand): There is no restriction for keeping the electricity usage within a limit in this scenario. Fig. 4 shows the daily cumulative cost comparison among policies for different devices, and Table V summarizes the results. The manually controlled policy has the maximum monthly total cost of $193. Among the rule-based approaches, the Pilloni *et al.* method [24] costs $166 and performs better than the Shirazi and Jadid method [13] with $180 monthly cost. The RL-based approach in [26] attains $172 monthly, and the proposed deep RL-based HERS policy achieves the lowest cost with $149 and minimizes the cost by 23% from the baseline manual control policy. The manually controlled policy starts operation immediately, thus does not take advantage of the lower electricity rate at off-peak hours, unlike the rule-based ones. However, the rule-based policy follows a conservative approach for optimization by searching low tariffs in a smaller time window to avoid creating resident discomfort. Especially, the EV charging time window in method [13] overlaps with the peak hours. So, these policies

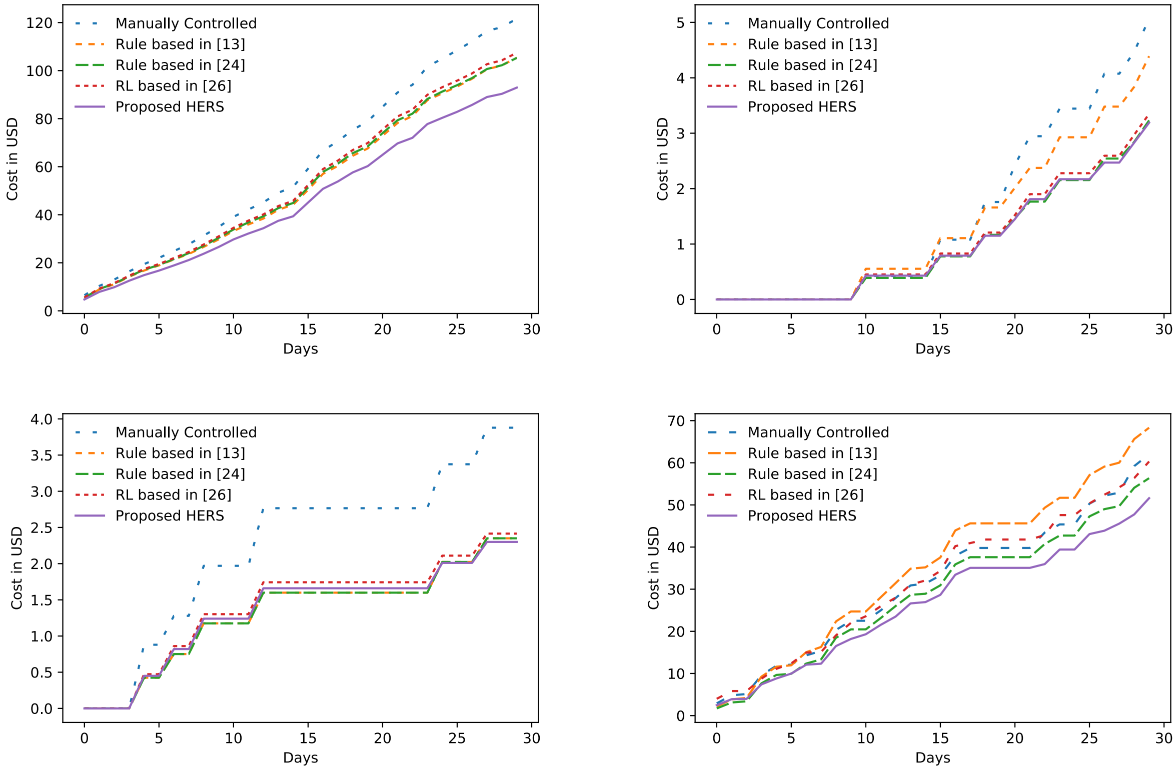


Fig. 4. Daily cumulative cost in Scenario 1 for devices: AC (top left), DW (top right), WD (bottom left), and EV (bottom right) for 1-month duration.

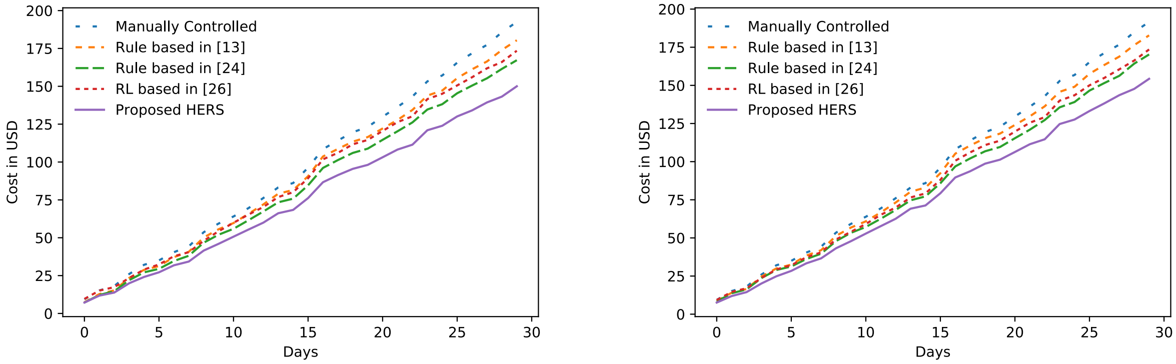


Fig. 5. Daily cumulative cost comparison for all devices among the considered policies for Scenario 1 (left) and Scenario 2 (right).

minimize the cost for all appliances on a smaller scale. The RL-based approach in [26] achieves comparable results with the rule-based policies. The success of this policy is limited due to only including electricity price in its state definition and overlooking many critical features that the HERS policy capitalizes on (see Table II). The HERS policy focuses on human feedback in its cost and runs the devices optimally. For instance, HERS keeps the AC off for shorter intervals during midnight without causing any resident discomfort. The proposed deep RL-based policy is expected to decrease the cost further for a system with more devices.

2) Scenario 2 (Limited Peak Demand): To avoid overloading a distribution system, the utility company often restricts users to keep energy usage under a threshold. Under this scenario, we limit the peak electricity usage to 10 kW to obey such restrictions. The EV charging can take up to 7.68 kW of electricity, even greater than the sum of other loads. So, all the devices other than the EV receive their unrestricted electricity. Hence, the other devices' electricity cost is the same for both scenarios. The EV charging gets the least priority and can consume up to the remaining electricity. Fig. 5 compares the total cost among different policies for both of the scenarios. In Scenario 2, all the policies attain similar results as in Scenario 1, however with a small increase in cost due to the restrictions. With more devices or lower peak limiting, the results may vary more compared with Scenario 1.

C. Computational Statistics

Fig. 6 shows that the proposed deep RL algorithm learns the optimal policy within 600 episodes. Table VI shows that training convergence takes 128 minutes and online decision making requires only 4 seconds in our computer (Intel Core i7,3.60 GHz, 16 GB RAM), exhibiting the real-world

TABLE VI
COMPUTATIONAL STATISTICS FOR THE EXPERIMENTS

Hardware	Software	Task	Computation time
Intel(R) Core i7,3.60	Python 3.7	Training	128 min
GHz, 16 GB RAM	Pytorch 1.8.1	Online Scheduling	4 sec

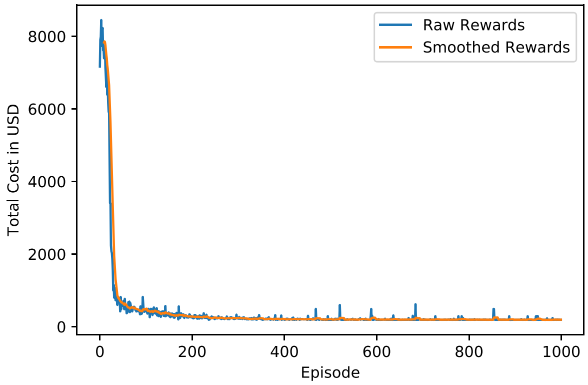


Fig. 6. Convergence of the proposed deep RL algorithm for HERS for scenario-1 total cost.

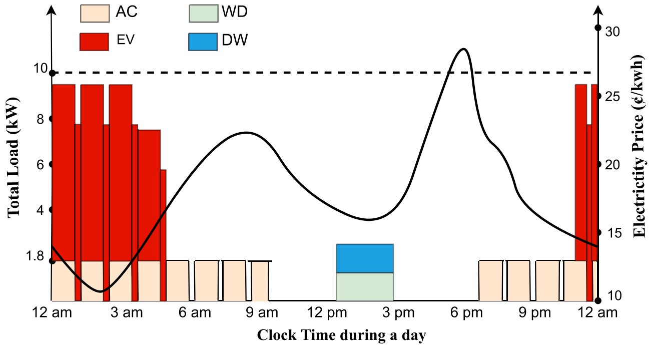


Fig. 7. HERS scheduling results for a day under scenario-2 (peak demand limit 10 kW). The black curve shows the electricity price.

applicability. The high cost in the early episodes indicates discomfort among residents; however, it ceases very fast. The trained HERS will take feedback from the consumer to optimize the electricity cost of the house. We examine the above policies for two scenarios.

D. HERS Schedule Demonstration

Fig. 7 shows the implemented schedule by HERS for a particular day. From the residents' feedback, HERS learns that switching the AC off for 15 minutes after 1 hour of continuous operation is its optimal schedule that minimizes electricity cost and does not create any discomfort. Hence, HERS follows this pattern and keeps the AC off when no one is home (9:30 am-5:45 pm). The EV is charged at maximum capacity (7.68 kW) during the low tariff early hours (12 am-3 am) and its remaining charge at 75% capacity (5.76 kW) during a slightly higher tariff (3 am-4 am). The EV returns home at 5:45 pm; however, it waits for lower electricity tariffs at 11 pm-12 am. The DW and WD require 2.5 hours of continuous power that the HERS schedules for the low-demand low-tariff hours during mid-day (12:30 pm-3 pm). Notably, HERS chooses this schedule instead of 12:00 pm-2:30 pm as the electricity price is lower during 2 pm-3 pm compared to 12 pm-1 pm. This sample schedule shows that HERS learns to minimize electricity cost and resident discomfort by utilizing the human feedback and activity labels in the proposed deep RL setup.

VI. DISCUSSION

This work focuses on key features derived from residents' activity for operating smart devices. The reward of the RL agent accommodates direct human feedback, thus providing a setup similar to the popular recommendation systems (e.g., video, book, music, etc.). We understand that any other approach incorporating more customized features for different appliances may achieve further improved results. So, the RL-based recommendation approach for device-specific policy-making has a high potential. This work demonstrates the benefit of including human activity-based states and human feedback-based rewards for adaptive HEM. Our model provides usage control of devices that do not include PV sources, energy storage, microgrid, and data sharing with other homes or a multi-agent setup. However, our core architecture can accommodate these features in the future to open up further research opportunities in this domain.

VII. CONCLUSION

This work presents a deep Reinforcement Learning (RL) based recommendation system for smart home energy management (HEM). Residents' manual override for a device is interpreted as a negative reward to the RL agent that operates the device. So, the goal of the RL agent is to capitalize low-tariff electricity without creating human discomfort. To the best of our knowledge, this is the first work that takes direct human feedback for device management in a general smart home setup. Intuitively, this method works similarly to the popular recommendation applications that suggest a video, book, music, etc., based on a user's usage pattern, so we call it Home Energy Recommendation System (HERS). Furthermore, the RL agent considers the human activities for state definition, another novelty the existing literature lacks. The experimental results show that the human activity pattern plays a vital role in device operation, in comparison with the RL approach of Xu *et al.* [26] that only considers electricity price for state definition. Our comparative analysis shows that HERS minimizes the electricity cost significantly with respect to the manually controlled policy, rule-based policies in [13], [24], and the RL-based policy presented in [26].

ACKNOWLEDGMENT

The authors would like to thank Dr. Hande Alemdar for providing and explaining the ARAS dataset.

REFERENCES

[1] F. Rocha *et al.*, "Energy efficiency in smart buildings: An IoT-based air conditioning control system," in *Proc. IFIP Int. Internet Things Conf.*, 2019, pp. 21–35.

[2] P. Siano, "Demand response and smart grids—A survey," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 461–478, Feb. 2014.

[3] H.-T. Roh and J.-W. Lee, "Residential demand response scheduling with multiclass appliances in the smart grid," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 94–104, Jan. 2016.

[4] B. Zhou *et al.*, "Smart home energy management systems: Concept, configurations, and scheduling strategies," *Renew. Sustain. Energy Rev.*, vol. 61, pp. 30–40, Aug. 2016.

[5] G. Strbac, "Demand side management: Benefits and challenges," *Energy Policy*, vol. 36, no. 12, pp. 4419–4426, 2008.

[6] F. Luo, G. Ranzi, S. Wang, and Z. Y. Dong, "Hierarchical energy management system for home microgrids," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5536–5546, Sep. 2019.

[7] Z. Zhao and C. Keerthisinghe, "A fast and optimal smart home energy management system: State-space approximate dynamic programming," *IEEE Access*, vol. 8, pp. 184151–184159, 2020.

[8] S. S. Shuvo and Y. Yilmaz, "CIBECS: Consumer input based electric vehicle charge scheduling for a residential home," in *Proc. North Amer. Power Symp. (NAPS)*, College Station, TX, USA, 2021, pp. 1–6.

[9] S. Sharma, Y. Xu, A. Verma, and B. K. Panigrahi, "Time-coordinated multienergy management of smart buildings under uncertainties," *IEEE Trans. Ind. Informat.*, vol. 15, no. 8, pp. 4788–4798, Aug. 2019.

[10] Y. Yang, J. Hao, Y. Zheng, and C. Yu, "Large-scale home energy management using entropy-based collective multiagent deep reinforcement learning framework," in *Proc. IJCAI*, 2019, pp. 630–636.

[11] W. Li, T. Logenthiran, and W. L. Woo, "Intelligent multi-agent system for smart home energy management," in *Proc. IEEE Innov. Smart Grid Technol. Asia (ISGT ASIA)*, Bangkok, Thailand, 2015, pp. 1–6.

[12] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Trans. Ind. Informat.*, vol. 7, no. 3, pp. 381–388, Aug. 2011.

[13] E. Shirazi and S. Jadid, "Optimal residential appliance scheduling under dynamic pricing scheme via HEMDAS," *Energy Build.*, vol. 93, pp. 40–49, Apr. 2015.

[14] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[15] P. N. Kolm and G. Ritter, "Modern perspectives on reinforcement learning in finance," *Mod. Perspect. Reinforcement Learn. Financ. J. Mach. Learn. Financ.*, vol. 1, no. 1, p. 28, 2020.

[16] S. S. Shuvo and Y. Yilmaz, "Predictive maintenance for increasing EV charging load in distribution power system," in *Proc. IEEE Int. Conf. Commun. Control Comput. Technol. Smart Grids (SmartGridComm)*, Tempe, AZ, USA, 2020, pp. 1–6.

[17] A. Haydari and Y. Yılmaz, "Deep reinforcement learning for intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 11–32, Jan. 2020.

[18] A. Nassar and Y. Yilmaz, "Reinforcement learning for adaptive resource allocation in fog ran for IoT with heterogeneous latency requirements," *IEEE Access*, vol. 7, pp. 128014–128025, 2019.

[19] S. S. Shuvo, Y. Yilmaz, A. Bush, and M. Hafen, "A Markov decision process model for socio-economic systems impacted by climate change," in *Proc. Int. Conf. Mach. Learn.*, 2020 pp. 8872–8883.

[20] S. S. Shuvo, M. R. Ahmed, H. Symum, and Y. Yilmaz, "Deep reinforcement learning based cost-benefit analysis for hospital capacity planning," in *Proc. Int. Joint Conf. Neural Netw.*, 2021, pp. 1–7.

[21] 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.

[22] H. Berlink and A. H. R. Costa, "Batch reinforcement learning for smart home energy management," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 2561–2567.

[23] W. Valladares *et al.*, "Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm," *Build. Environ.*, vol. 155, pp. 105–117, May 2019.

[24] V. Pilloni, A. Floris, A. Meloni, and L. Atzori, "Smart home energy management including renewable sources: A QoE-driven approach," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2006–2018, May 2018.

[25] F. Marcello and V. Pilloni, "Smart building energy and comfort management based on sensor activity recognition and prediction," *Sensors*, vol. 1, p. s2, 2020.

[26] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A multi-agent reinforcement learning-based data-driven method for home energy management," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3201–3211, Jul. 2020.

[27] M. Khan, J. Seo, and D. Kim, "Real-time scheduling of operational time for smart home appliances based on reinforcement learning," *IEEE Access*, vol. 8, pp. 116520–116534, 2020.

[28] L. Yu, T. Jiang, and Y. Zou, "Online energy management for a sustainable smart home with an HVAC load and random occupancy," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1646–1659, Mar. 2019.

[29] S. Bahrami, V. W. S. Wong, and J. Huang, "An online learning algorithm for demand response in smart grid," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4712–4725, Sep. 2018.

[30] J. Y. Park, T. Dougherty, H. Fritz, and Z. Nagy, "*LightLearn*: An adaptive and occupant centered controller for lighting based on reinforcement learning," *Build. Environ.*, vol. 147, pp. 397–414, Jan. 2019.

[31] D. Liciotti, M. Bernardini, L. Romeo, and E. Frontoni, "A sequential deep learning application for recognising human activities in smart homes," *Neurocomputing*, vol. 396, pp. 501–513, Jul. 2020.

[32] D. J. Cook, M. Schmitter-Edgecombe, L. Jönsson, and A. V. Morant, "Technology-enabled assessment of functional health," *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 319–332, 2018.

[33] C. Chen, D. J. Cook, and A. S. Crandall, "The user side of sustainability: Modeling behavior and energy usage in the home," *Pervasive Mobile Comput.*, vol. 9, no. 1, pp. 161–175, 2013.

[34] A. Sanchez-Comas, K. Synnes, and J. Hallberg, "Hardware for recognition of human activities: A review of smart home and AAL related technologies," *Sensors*, vol. 20, no. 15, p. 4227, 2020.

[35] M. Aktukmak, Y. Yilmaz, and I. Uysal, "A probabilistic framework to incorporate mixed-data type features: Matrix factorization with multimodal side information," *Neurocomputing*, vol. 367, pp. 164–175, Nov. 2019.

[36] M. Chen, A. Beutel, P. Covington, S. Jain, F. Belletti, and E. H. Chi, "Top-k off-policy correction for a reinforce recommender system," in *Proc. 12th ACM Int. Conf. Web Search Data Min.*, 2019, pp. 456–464.

[37] P.-H. Su, P. Budzianowski, S. Ultes, M. Gasic, and S. Young, "Sample-efficient actor-critic reinforcement learning with supervised data for dialogue management," 2017, *arXiv:1707.00130*.

[38] "NYISO Real-Time Dashboard." NYISO. 2022. [Online]. Available: https://www.nyiso.com/real-time-dashboard

[39] H. Alemdar, H. Ertan, O. D. Incel, and C. Ersoy, "ARAS human activity datasets in multiple homes with multiple residents," in *Proc. 7th Int. Conf. Pervasive Comput. Technol. Healthcare Workshops*, Venice, Italy, 2013, pp. 232–235.

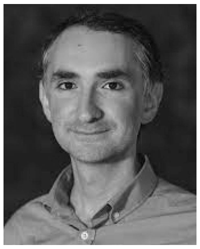
[40] A. Burdick, "Strategy guideline: Accurate heating and cooling load calculations," Nat. Renew. Energy Lab.(NREL), Golden, CO, USA, Rep. NREL/SR-5500-51603, 2011.

[41] P. Berkholz, R. Stamminger, G. Wnuk, J. Owens, and S. Bernarde, "Manual dishwashing habits: An empirical analysis of U.K. consumers," *Int. J. Consum. Stud.*, vol. 34, no. 2, pp. 235–242, 2010.

[42] "Bosch 500 Series- Stainless Steelshp65T55UC Instruction Manual." Bosch. [Online]. Available: https://media3.bosch-home.com/Documents/9001218494_A.pdf (Accessed: Apr. 26, 2021).

[43] M. Muratori, M. J. Moran, E. Serra, and G. Rizzoni, "Highly-resolved modeling of personal transportation energy consumption in the United States," *Energy*, vol. 58, pp. 168–177, Sep. 2013.

Salman Sadiq Shuvo (Graduate Student Member, IEEE) received the M.Sc. degree in electrical engineering from the University of South Florida, where he is currently pursuing the Ph.D. degree. His research interests are in reinforcement learning based optimization methods that he has employed in several environmental and energy systems projects.



Yasin Yilmaz (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Columbia University, New York, NY, USA, in 2014. He is currently an Assistant Professor of Electrical Engineering with the University of South Florida, Tampa. His research interests include statistical signal processing, machine learning, and their applications to computer vision, cybersecurity, IoT networks, energy systems, transportation systems, and communication systems.