To Heat Or Not To Heat: Reinforcement Learning for Optimal Residential Water Heater Control

Hallie Dunham
Dept. of Electrical Engineering
Stanford University
hdunham@stanford.edu

Eun Seo Jo
Dept. of Computer Science
Stanford University
eunseo@stanford.edu

Aditya Khandelwal
Dept. of Computer Science
Stanford University
akhand@stanford.edu

Abstract

This paper presents a reinforcement learning based approach for determining the optimal control strategy of a residential water heater. The objective is to minimize electricity costs while meeting temperature requirements when hot water is needed. Following an overview of prior efforts at solving this problem, we outline the problem statement for this research paper, simulate the behavior of a residential water heating appliance, and implement a model-free, online solution to arrive at an optimal policy. Additionally, in order to realistically model the problem, we refer to real-world water heater usage data, including a dataset with circuit-level power consumption data for a community of residential households in Texas. Finally, we provide an explanation of the results of our evaluation of the generated optimal policy and expound on our search for a reward system that fairly distributes rewards and penalties so that our algorithm can arrive at the correct set of solutions.

1 Introduction

Numerous government, public, and private institutions undertake the task of modeling energy demands in order to predict the amount of supply needed in the future. The first wave of innovation in the way we consume energy supplied through grid-lines came with the advent of pricing models that deterministically charge consumers more or less based on some average supply heuristics derived from previous energy consumption data. However, with recent advancements in renewable energy technologies and the proliferation of micro-grids, it is becoming increasingly difficult for traditional players to model their energy supply needs. This means that a dynamic & probabilistic model of pricing energy is on the horizon. Consequently, this also implies that people will find it difficult to reduce their electricity bills by simply switching off their devices during peak hours. This presents an even more pressing issue for industrial players that have many unoptimized heating units that should ideally be used only during off-peak hours or when they are needed the most.

1.1 Problem Definition

In this paper, we hypothesize that the aforementioned problem can be simplified to a single machine that is used to heat water, where the decision to consider is whether to use the machine based on the 1. time of the day, 2. the variable pricing of electricity and 3. the prior temperature of the water. Moreover, we factor in 4. noise because of outside temperature and the usage of water by people living in the residence in which this water heating appliance is installed. Our key objective is to find an optimal action (to turn the heater on or off) to take at each temperature observation for a given time.

Since this is a decision-making problem, it is appropriate to model this scenario as a Markov Decision Process (MDP). An MDP is a type of sequential decision process whereby an agent can choose from a

CS238: Decision Making Under Uncertainity, Fall 2019, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

set of actions at each time-step after observing their state at the current time. It is fundamentally based on the Markov assumption, which states that the next state in a system only depends on the current state and the action chosen by the agent. There are a variety of solution methods for MDPs, however, the one that we have implemented here is SARSA-λ, which is an online, model-free algorithm.

2 Related work

Water heating optimization is a simplified version of the problem of electricity spending optimization that researchers have worked on using various methods.

One approach is using dynamic programming to model heating schedules (1). This project uses weather forecasts, hot water consumption, energy price, and electrical demand to find optimal ways of heating a water tank with minimal cost. The authors show promising results for a stochastic DP approach to control the water heater in keeping the minimum hot water needed in a tank. Recently, research has used an MPC (model predictive control) (2) approach to model household device usage to reduce real time pricing for residential environments. Previous research explored predictive load scheduling as a way to reduce electricity costs (3). In this paper, the trade-off between the cost of electricity and conservatism degree is also discussed. Another approach is based on Djikstra's algorithm to schedule optimal loads to heating appliances (4). This paper proposes a more flexible way of building a system that can be fine-tuned to the heating preferences of each individual resident of a household. Finally, the Monte Carlo method is used to simulate power consumption of multiple residential water heaters (5). In this paper, a temperature state priority list is managed in order to reduce peak demand for the micro-grid as a whole.

Based on this work, we decided to model the problem as an MDP, as described above, and use a reinforcement learning based approach to find an optimal scheduling policy that is fairly flexible in a variety of different settings.

3 Dataset & Inspiration

We were inspired to work on this problem through a dataset provided by the Dataport at Pecan Street. This API lets us query real-world information on household level data of appliance usage based on time and date for a preset range of days. We analyzed this dataset to understand how to model the usage of a single appliance within a single household. This dataset helped us form our intuition of the appropriate reward weights for any potential state-action pair and that time of day is a determining factor of heat demand. The dataset also inspires us to see how these models can be taken forward and applied to more than just households within a community, but to industries and public areas as well.

4 Methods

We formulate the task of controlling a water heater as a model-free MDP. The state space, action space, and rewards are described in detail below. We built a simulator model to generate sample data to run SARSA-λ.

4.1 State Space

We modeled our state as a tuple of two values, namely, temperature of water in the heating unit T and time of the day t. Since both temperature and time have an infinite and somewhat continuous range, we decided to discretize these observations. Therefore, for our simplified version of the original problem, temperature T is divided into a 110 potential values ranging from $T_{\text{min}} = 40°$ F to $T_{\text{max}} = 150°$ F at intervals of $1°$ F, which is a fair assumption to make because we do not expect the water in a water heater to be freezing cold or scalding hot. For time t, our discretized values range from 0 to 23 which represents hours of a day. Therefore, our state space, S, consists of 2664 unique states that a water heating system can exist in.

4.2 Action Space

Our action space consists of only two actions - turning the heater off (represented as a 0) and turning the heater on (represented as a 1). Therefore, the size of our action space, A, is 2. Consequently, this also means that the size of our state-action space is 5328.

4.3 Transitions

The transition function is defined as a function that results in an incremental state from a previous state based on a certain action. Time always deterministically increases by one (wrapping back to 0 after 23). However, the temperature of water is probabilistically computed. We do it by sampling temperature from a probability distribution determined by the current temperature, the expected usage at a given time of day, the outside temperature at a given time of day and whether or not the water was being already heated. We use a scaled and shifted sample from a beta distribution with $\alpha = 4$ and $\beta = 2$. Therefore, our transition function mandates that the range of temperatures remain between T_{lb} and T_{ub}, calculated as follows:

$$T_{\text{lb}} = \min(T_{\text{max}}, \max(T_{\text{min}}, T + a \times T_{\text{heated}} - T_{\text{maxdrop}}))$$
$$T_{\text{ub}} = \max(T_{\text{min}}, \min(T_{\text{max}}, T + a \times T_{\text{heated}}))$$

T_{maxdrop} is the maximum drop in temperature due to water usage for a given period of time. When hot water is used, cold water replaces it in the tank and the temperature drops. The value of T_{maxdrop} is a function of the difference between the current water temperature and the new cold water temperature. We estimated this function using a standard tank volume of 80 gallons, an average household hot water usage of 45 gallons/day, and the assumption that hot water is likely to be used the most between 5:00 AM to 9:00 AM and 5:00 PM to 1:00 AM. We assume that having the heater on deterministically increases the water temperature by $T_{\text{heated}} = 23^\circ$ F. This value is calculated using a standard residential water heater power of 4 kW, the same tank volume as above, and the specific heat of water, as described in the following equations.

$$4.5 \text{ kW} \times 1 \text{ hour} = 4.5 \text{ kWh} \tag{1}$$

$$80 \text{ gal} \times 3.79 \text{ kg gal}^{-1} \approx 300 \text{ kg} \tag{2}$$

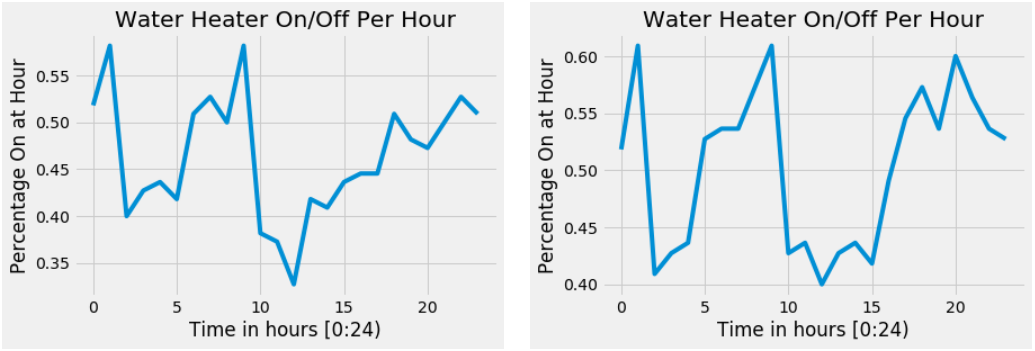
$$300 \frac{\text{kcal}}{\Delta^\circ\text{C}} \times 1.16 \times 10^{-3} \frac{\text{kWh}}{\text{kcal}} = 0.349 \frac{\text{kWh}}{^\circ\text{C}} \tag{3}$$

$$\frac{4.5 \text{ kWh}}{0.349 \frac{\text{kWh}}{\Delta^\circ\text{C}}} \approx 12.9\Delta^\circ \text{ C} \tag{4}$$

$$T_{\text{heated}} = 12.9\Delta^\circ \text{ C} \times 1.6 \frac{\Delta^\circ \text{ F}}{\Delta^\circ \text{ C}} \approx 23\Delta^\circ \text{ F} \tag{5}$$

4.4 Reward Model

Our model of the reward function is a piece-wise combination of two components: the cost of electricity and the boolean notion of the water temperature not being within a desired bound when residents may want to use it. The latter is modeled into the reward function as a penalty, while the former depends on the time of the day and the action being executed. Therefore, electricity cost is a function of time and action. We assume a standard time-of-use pricing scheme with a peak pricing period from 4:00 PM to 9:00 PM. The off-peak price is 0.10 $/kWh and the peak price is 0.20 $/kWh. As calculated above, the heater uses 4.5 kWh of energy if it is switched on for an hour. The cost of taking action $a = 1$, i.e., switching the heater on, is the product of the price of electricity and the energy used. We use desired temperature bounds of 100° F and 140° F. If the temperature is not within these bounds between 5:00 AM to 9:00 AM and 5:00 PM to 1:00 AM, then a penalty of 3 is incurred. All these constants defining our reward model can be very easily modified to match a the specific electricity rate structure, temperature preference, and prioritization of cost versus temperature



(a) Reward system A

(b) Reward system B

Figure 1: Heater on/off throughout hours in the day

precision for a household. This makes our approach flexible enough to fit other energy optimization tasks that require an individualized optimal control policy for a different use case (such as a different appliance or group of appliances).

4.5 Learning Algorithm

We use SARSA-λ to perform reinforcement learning on our simulated data to find the optimal control policy for the water heating appliance. This algorithm is very similar to other reinforcement learning algorithms, such as Q-Learning, except that it is online and learns the Q-value based on the action performed by the current policy instead of the greedy policy (6). The update for the value iteration used is as follows:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$

where α is the learning rate of the algorithm and γ is a hyper-parameter that can be tuned according to the training data.

5 Results & Evaluation

We analyze our final generated policies to understand how our algorithm performed on the state space given different rewards. We also tuned hyperparameters and provide explanation on how an optimal policy for this problem was arrived at by our algorithm.

We use two rewards systems. In reward system A, we assign a one constant reward for water being in the optimal range at the right time in the day. Reward system B uses a piecewise model where the reward $r_1 < r_2$ is given for being in range 100-120 °F and reward r_2 is given for being in range 120-140 °F. In Figures 1a and 1b we plot the policy results from SARSA-λ from both rewards systems. In both cases we see that at hours of the day where people are home and awake to use water, 5:00 AM to 9:00 AM and 5:00 PM to 1:00 AM, the heater is more likely to be on than off. The peaks are sharper in 1b because the piecewise model incentives the algorithm to pursue the a narrower range of water temperature, to optimize reward. We ran these experiments separately to try out a more fine-grained system.

6 Conclusion & Future Work

In conclusion, SARSA-λ was able to learn optimal times of the day to turn on the water heater which seem reasonable intuitively. And in fact, when the reward system is stepwise, the results are more clear-cut. These findings show that an MDP is a good approach for modeling stochastic pricing for household electricity usage.

In future work, we can add in other factors that make the model more realistic. One is to simulate the the gradual fall in water temperature over time. Even if the hot water is in a heat-conserving tank, the gradual natural fall in temperature is inevitable. Another factor is to simulate hot water consumption. As more heated water is consumed from the tank, more heating will be needed to maintain the temperature range. There are of course other ways to make the model more realistic by taking weather data, household water consumption patterns, and electricity pricing from the real-world.

7 Contributions

All members contributed equally to the paper write-up. We have worked on other aspects of the project individually or in subgroups. Aditya wrote and ran the code for SARSA-λ. He also wrote and ran experiments to tune parameters on the model and generate result charts. Eun Seo and Hallie conceived the project idea and wrote the code to simulate the environment. They ran the simulation to generate sample data to run the policy iteration algorithms.

References

[1] C. Passenberg, D. Meyer, J. Feldmaier, H. Shen, (2016 IEEE International Energy Conference). Optimal water heater control in smart home environments. *doi: 10.1109/ENERGYCON.2016.7513964*

[2] Wang, Jidong et al. "MPC-Based Interval Number Optimization For Electric Water Heater Scheduling In Uncertain Environments". Frontiers In Energy, 2019. Springer Science And Business Media LLC, doi:10.1007/s11708-019-0644-9.

[3] Wang, Jidong et al. "A Robust Optimization Strategy For Domestic Electric Water Heater Load Scheduling Under Uncertainties". Applied Sciences, vol 7, no. 11, 2017, p. 1136. MDPI AG, doi:10.3390/app7111136.

[4] Kapsalis, Vassilis, and Loukas Hadellis. "Optimal Operation Scheduling Of Electric Water Heaters Under Dynamic Pricing". Sustainable Cities And Society, vol 31, 2017, pp. 109-121. Elsevier BV, doi:10.1016/j.scs.2017.02.013.

[5] Yin, Zhaojing et al. "Optimal Scheduling Strategy For Domestic Electric Water Heaters Based On The Temperature State Priority List". Energies, vol 10, no. 9, 2017, p. 1425. MDPI AG, doi:10.3390/en10091425.

[6] Sutton. R.S., Barto A.G. (2018). Reinforcement Learning: An Introduction (chapter 6.4). *http://incompleteideas.net/book/ebook/node64.html*