Autonomous Thermal Control System

Omid Ardakanian
David R. Cheriton School of Computer Science
University of Waterloo
Technical Report CS-2010-16

Abstract

Building heating and cooling accounts for nearly 18-24% of all energy usage. Building energy management can reduce both its operating costs and its carbon footprint. We propose the use of a factored Partially Observable Markov Decision Process (POMDP) to model and efficiently control a commercial building heating system. We use supervised learning and ground truth data to capture the parameters of the POMDP. We show that our solution outperforms the HVAC system in terms of energy efficiency.

1. Introduction

Energy consumption of residential and commercial buildings currently accounts for about 30-40% of global energy use (UNEP 2007). About 60% of this energy is used for space heating (or cooling), 18% for water heating, and only 3% is used for lighting. In high- and middle-income countries, energy is mostly generated from fossil fuel, directly contributing to global climate change. Our work, therefore, uses decision making techniques to reduce inefficiencies in space heating and cooling. Specifically, we use decision making to control the temperature of a building according to the activity pattern of its occupants. We believe that planning is useful in optimizing the thermal control problem because it takes time to warm up a house.

2. Related Work

Energy resource monitoring and activity recognition have been extensively studied in the literature. The SpotLight system (Kim et al. 2008) monitors a user's energy usage profile using wireless sensors, assuming that user proximity is the cause of measured energy usage. The ACme (Jiang et al. 2009) system monitors household power consumption in real time at a power outlet to help occupants understand their electricity usage pattern. It does not monitor user activity. The HydroSense (Froehlich et al. 2009) system, similarly, monitors home water usage. Finally, the ViridiScope (Kim et al. 2009) project monitors power usage indirectly by sensing signals emitted by an electrical appliance. Our work differs from these projects in two respects. First, we study space heating and cooling, which account the bulk of energy usage in most buildings. Second, we use an automated decision making system to optimize thermal control.

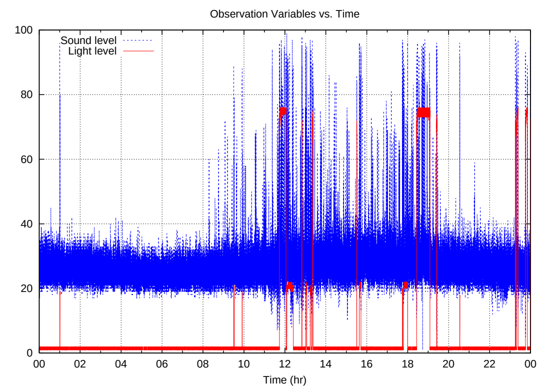


Figure 1: Sound and light measurement for a day

3. Overview

We propose a fully autonomous decision-making system to optimally control a building heating system. Our main goal is to not expend energy on heating or cooling a space in the building if there is no occupant currently present; although, in the initial model we only tackle the space heating problem. We say that a room in a building is *active* if it has an occupant. To preserve privacy, we do not directly observe if a room is occupied. Instead, we use measurements of sound and light levels to infer activity. For instance, Figure 1 shows the sound and light level we measured in a typical room over the period from midnight to midnight on a typical day. We see that sound levels increase perceptibly at around 9am and drop off at 7:30pm, resuming at around 11:30pm. Light levels also show a similar trend. We verified that these correspond to room activity by students and by custodial staff. This suggests that a control system based on these measurements would make good decisions.

We define a *policy* to be the sequence of decisions that control the temperature of a room in a building based on the control system's belief about current and past activity patterns. We use a POMDP to find optimal thermal control policies of a single room. This is a substitute for the traditional thermostat devices, which are oblivious to activity, and therefore likely to be ineffective in practice.

3.1 Model

Figure 2 is the model of the system. The state variables are the activity in a room (S^A), its temperature (S^T), and the half-hour interval of time within a day (S^C). To keep the model tractable, temperature and clock state variables are discretized. In our model, the domains of these state variables are defined as follows:

$Dom(S^A) = \{Active, Inactive\}$
$Dom(S^T) = \{10, 10.5, 11, ..., 29, 29.5, 30\}$
$Dom(S^C) = \{0:30, 1:00, ..., 23:30, 24:00\}$

The belief of being in a certain activity state is updated based on the value of two $observation$ $variables$: sound and light levels, denoted by O^S and O^L respectively. Although our sensors can measure sound and light levels with high precision, we find it sufficient to define only three values for sound level: *high, normal and low* and two values for light level: *On and Off*.

We define only two possible actions A: *blocking the heating vent* and *unblocking the vent*. Blocking a vent decreases energy consumption and is assumed to change the temperature by a value δ (corresponding to heat loss or gain) in one time step. Symmetrically, unblocking the vent increases energy consumption and is assumed to change the temperature by δ' in one time step. Given fixed parameters such as the size of the room, the difference between indoor and outdoor temperatures, and the capacity of the heating system, we can compute δ and δ' from analytic models. To this end, we use the monthly average outdoor temperature to compute δ and δ' for possible values of the indoor temperature.

Finally the reward function R is defined as a function both of the comfort of the occupants and of the energy consumed. If a room is active, then the reward is $|T_{preferred} - T| \times C$, where $T_{preferred}$ is the preferred temperature of the room, T is the deterministically measured temperature, and C is the weight of user's satisfaction. If a room is inactive, the reward is $|T - T_{setpoint}| \times C'$ where $T_{setpoint}$ is the minimum temperature of the room and C' is the weight of energy saving. Since heating increases the temperature of the room, optimizing the energy consumed for heating is equivalent to finding the optimized temperature profile.

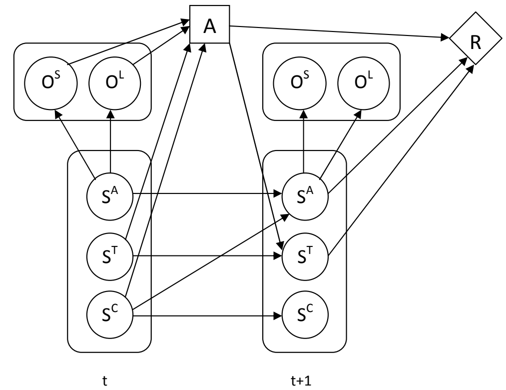


Figure 2: POMDP model

Determining the values of δ, δ', C, C', $T_{preferred}$ and $T_{setpoint}$ is discussed in section 3.3.

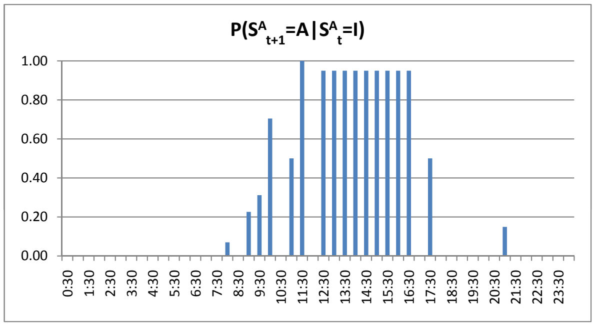
3.2 Learning Model Parameters

To learn standard transition and the observation functions of the POMDP, we have deployed 24 WeatherDuck v2 sensors measuring the ambient temperature, humidity, light level, sound level, and air flow. Sensors are attached to serial ports of a subset of 40 Linux-based embedded systems deployed in offices, labs and public areas in the Davis Center at the University of Waterloo (Ahmed and Ismail 2009). We poll sensors every two seconds and collect data from these sensors into a controller weekly. To gather ground truth data, we have recorded occupancy of one of the monitored labs using a log sheet; filled out voluntarily by the lab members.

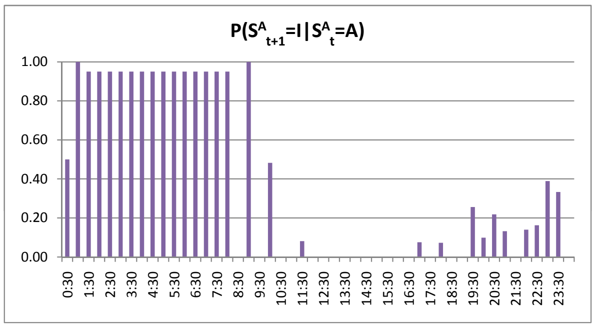
Using the occupancy record of 15 days, we have computed the transition probability for the activity state variable, S^A. This is the probability of transition to an active/inactive state given that the previous clock state was T and we have been in active/inactive state. So the transition probability will be described by these probabilities at different time slices during a day:

$P(S^A_{t+1} = Active \mid S^A_t = Inactive, S^C_t = T)$
$P(S^A_{t+1} = Inactive \mid S^A_t = Active, S^C_t = T)$
$P(S^A_{t+1} = Active \mid S^A_t = Active, S^C_t = T)$
$P(S^A_{t+1} = Inactive \mid S^A_t = Inactive, S^C_t = T)$

Figure 3 shows the first two probabilities over time. Since the prior probability is zero at some time slices (for instance, during the night) we assigned reasonable values to the transition probability during these periods. More specifically, we used 0.05 for staying in the active state and 0.95 for the transition to the inactive state from the active state during the night.



(a) Probability that the state of the room is changed to active from inactive



(b) Probability that the state of the room is changed to inactive from active

Figure 3: Learning the transition function

We have computed the observation function by integrating the sensor measurement and the activity data recorded by using the log sheet. Then we marginalized out the clock state and computed probabilities as shown in Table 1.

3.3 Variations of the Problem

To see the effect of the temperature change rate and user's satisfaction and energy saving weights on the optimal policy, we have defined four different problems. Table 2 summarizes configuration of these problems. Different values are used for the weight of user's satisfaction , C, the weight of energy saving, C', and the temperature change in half an hour, δ and δ'.

Values of δ and δ' are experimentally determined by measuring the temperature change in an office at the University of Waterloo during February 2010 when the average daily temperature was reported $-2°$C and the indoor temperature was 21°C. These values are used in CICA10 and CI10CA25 problems where we assumed that δ and δ' are fixed for all indoor temperatures. However, in CICA10MD and CI10CA25MD problems these values are assumed to be a function of the difference between indoor and outdoor temperatures. We set the value of $T_{preferred}$ to be 23°C according to the ASHRAE standard 55 (ASHRAE 2004). We also set the value of $T_{setpoint}$ to be 15°C.

On and High given Active	0.00944
On and Medium given Active	0.41881
On and low given Active	0.57170
Off and High given Active	0.00000
Off and Medium given Active	0.00003
Off and low given Active	0.00002
On and High given Inactive	0.00014
On and Medium given Inactive	0.00692
On and low given Inactive	0.00978
Off and High given Inactive	0.00031
Off and Medium given Inactive	0.27339
Off and low given Inactive	0.70946

Table 1: The observation function

Problem	C	C'	δ	δ'
CICA10	10	10	-0.5	1
CI10CA25	25	10	-0.5	1
CICA10MD	10	10	-1 if $S^T \geq 21$ -0.5 otherwise	1 if $S^T \geq 21$ 1.5 otherwise
CI10CA25MD	25	10	-1 if $S^T \geq 21$ -0.5 otherwise	1 if $S^T \geq 21$ 1.5 otherwise

Table 2: Configuration of different variations of the problem

4. Experimental Results

In this section we briefly explain three different algorithms that we have used to solve the POMDP with 3936 states. Then we compare optimal value functions of these algorithms with the upper bound value function (MDP value function).

To solve the factored POMDP using the Perseus algorithm, we have used the Symbolic Perseus package (Poupart). This package solves POMDPs using the symbolic version of the Perseus algorithm which uses the Algebraic Decision Diagram (ADD) for compact representation.

Additionally, we made use of the ZMDP package (Smith) to solve the equivalent flat POMDP using two heuristic value iteration algorithms. More specifically, we have used the FRTDP and HSVI2 algorithms for solving the thermal control problem.

4.1 Focused Real-Time Dynamic Programming

Focused Real-Time Dynamic Programming (FRTDP) (Smith and Simmons 2006) is an asynchronous value iteration algorithm that updates most relevant states more often. In this algorithm, the selected node for value update is the one that is visited in the forward exploration of the search tree according to a heuristic. In other words, the graph search algorithm for finding the optimal policy is limited to a subset of the search graph, called the *explicit* graph. Then the *explicit* graph is extended by adding some nodes that have higher relevance to the policy. The relevance of a state is defined by the priority value of that state. This value represents the benefit of directing the search to that state. In FRTDP, a node with the largest contribution to the uncertainty of the optimal value function has the highest priority.

To find the optimal value function, two admissible bounds (the upper and lower bounds) for the value function are used. Therefore, the approximate value function can be found by squeezing these bounds. The value iteration algorithm is terminated when the difference between the higher bound and the lower bound is less than a value ϵ or when the search depth is higher than a maximum depth. FRTDP maintains the maximum depth adaptively.

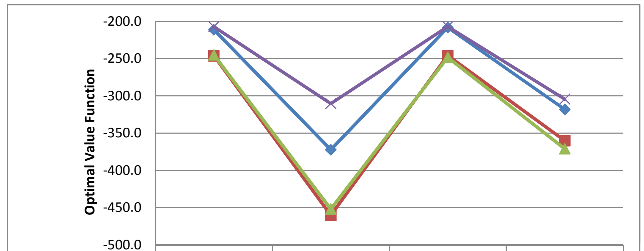
FRTDP's convergence is guaranteed under a set of conditions. It is shown that FRTDP converges faster than other heuristic value iteration algorithms for finding optimal policies in MDP.

4.2 Heuristic search value iteration algorithm

Computing the exact value functions in different iterations of the Bellman's equation is not practical for large size problems. Moreover, it is practically observed that performing fast approximate updates often results better than few exact updates. This is the intuition behind updating the value function approximately. Point-based value iteration algorithms generate a single α vector that is maximal at a belief point b.

Heuristic search value iteration algorithm (HSVI) (Smith and Simmons 2004) is a point-based value iteration algorithm that maintains both upper and lower bounds on the optimal value function and selects the action and observation heuristically. According to the IE-MAX heuristic, an action with the largest upper bound is selected. The observation selection is enhanced by the weighted excess uncertainty heuristic. This heuristic helps to focus on the child node which has the greatest contribution to excess uncertainty at the parent.

HSVI makes a local update at a specific belief. These beliefs are chosen by exploring the search tree according to the aforementioned heuristics. The upper bound is updated by adding a point to the point set and applying the max-project operator on the point set. The lower bound update is done as before by adding a vector. HSVI2 (Smith and Simmons 2005) is the latest implementation of HSVI which uses tighter initial bounds, avoids solving linear programs and makes better use of sparsity.



	CICA10	CI10CA25	CICA10MD	CI10CA25MD
Perseus+ADD	-211.3	-372.3	-208.4	-318.2
FRTDP	-246.4	-460.3	-245.9	-360.2
HSVI	-245.1	-451.8	-248.3	-371.0
MDP	-207.3	-310.4	-207.1	-304.4

Figure 4: Optimal value function of different problems using different algorithms

4.3 Symbolic Perseus and the ADD representation

Perseus (Spaan and Vlassis 2005) is a point based value iteration algorithm. The value iteration is performed by doing partial point-based backups at reachable belief states. A point based backup computes the best α-vector for each point in a set of witness belief states.

The Algebraic Decision Diagram representation exploits the context-specific independence to compactly represent a function defined over a set of variables. Therefore, representing the transition, observation and reward functions using ADDs, results in an efficient implementation of matrix operations. To tackle large scale POMDPs, we can integrate the Perseus algorithm with ADD representation (Poupart 2005).

4.4 Computing the Optimal Value Function

Figure 4 represents the average expected reward for 500 runs. We did not plot the run-time of these algorithms because it is almost the same for all of them. Moreover, the POMDP should be solved once a month in an off-line fashion. Therefore, we do not have any time constraints.

The Perseus algorithm seems to be closer to the upper bound of the value function (the MDP value function.) Furthermore, value functions of the FRTDP and HSVI2 algorithms are similar in all variations of the problem.

4.5 Tracing an Optimal Policy

To have a better sense of the efficiency of optimal policies, we plotted the 24-hour temperature change that is the outcome of performing a sequence of actions selected by the Perseus+ADD algorithm. To this end, we traced optimal policies that are found by this algorithm. We continued tracing the optimal policy until we found the steady state; where the initial temperature is equal

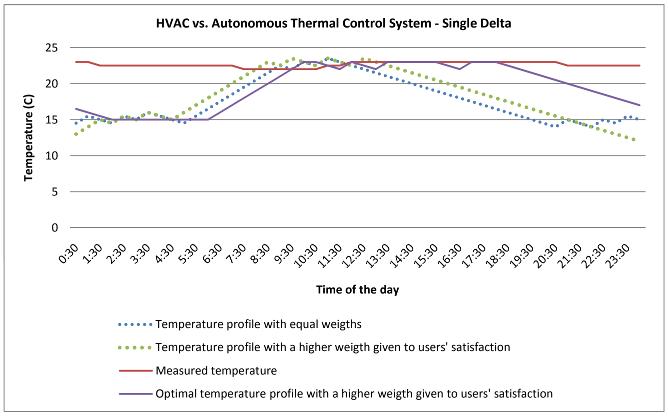


Figure 5: Comparison of fixed rate temperature change problems with the optimal and measured temperature profiles

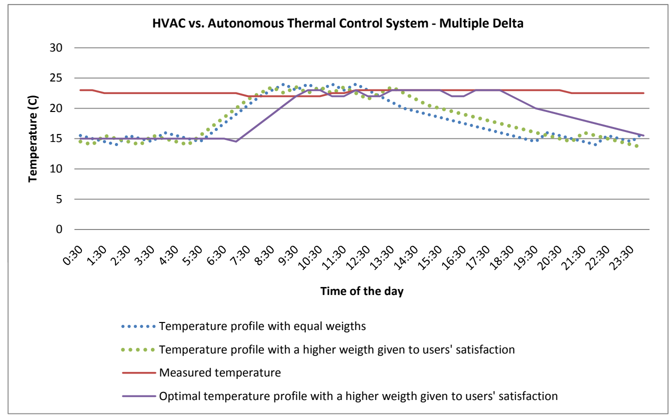


Figure 6: Comparison of variable rate temperature change problems with the optimal and measured temperature profiles

to the final temperature. We also plotted the 24-hour temperature change of the optimal thermal control policy as well as the monitored temperature of the room. The monitored temperature shows that the HVAC system keeps the temperature almost stable regardless of the occupancy of the room. The optimal thermal control policy is the most energy efficient temperature profile for the recorded activity data in the case where we assign the highest weight to user's satisfaction when the room is occupied and the highest weight to energy saving when it is empty.

Figure 5 shows the 24-hour temperature change of the optimal policies of CICA10 and CI10CA25 problems. Similarly, Figure 6 shows the 24-hour temperature change of optimal policies of variable rate temperature change problems; CICA10MD and CI10CA25MD. Increasing the weight of user's satisfaction, our system starts warming up the room earlier than the case with equal weights. It also stops warming up the room (by blocking the vent) later than the other case. This is exactly consistent with our expectation.

It should be mentioned that we optimize the temperature instead of energy consumption of the heating system. However, it is possible to estimate energy consumption from the operation time and nameplate power of the heating system. Specifically, it is equal to the duration of the time that the temperature curve has positive slope (assuming that the outdoor temperature is higher than the indoor temperature.) With this assumption, it turns out that in variable rate temperature change problems; CICA10MD and CI10CA25MD, energy consumption of the proposed system is equal to 62.5% of the HVAC system's energy consumption. Therefore, the widespread implementation of our solution would greatly reduce carbon emissions.

5. Future Work

Extending this work to the cooling system is one of the future directions. To this end, the reward model should be changed to be a function of energy consumption rather than temperature. Based on the result of this work, we plan to refine the model. We want to study sensitivity of our results to the level of discretization, the length of time steps and the setpoint temperature.

Another possible extension is to learn the weekly schedule instead of the daily schedule. Since occupancy of a building do relate to the day of week as well as the time of the day, learning the weekly schedule would increase the accuracy of our prediction and the optimality of the selected policy.

We are also interested in computing the total cost of heating a room up to a temperature under different configurations of the problem. We would also like to find the answer of the following question: what should be the ratio of user's satisfaction weight to energy saving weight in limited budget scenario?

Moreover, it is known that the standard temperature of a room depends on the culturally-induced clothing norms of its occupants (Fountain, Brager, and de Dear 1996). Therefore, borrowing some ideas from the Multi-agents context, we plan to find dominant strategies of the occupants of a building (agents) when they can set the temperature of the building and also change the clothing condition. It is also interesting to analyze the game where goals of government energy agencies, power generators and householders are in contrast. Finding Nash Equilibria of this game is of high interest too.

6. Conclusion

We presented the use of POMDPs to model and solve a real world problem; measurement-based building thermal control. We believe that the use of a sophisticated decision-making approach combined with a large-scale sensor deployment and field measurements will allow us to moderate energy consumption of the heating system.

References

Ahmed, N., and Ismail, U. 2009. Designing a High Performance WLAN Testbed for Centralized Control. *Testbeds and Research Infrastructures for the Development of Networks and Communities, International Conference on* 1–6.

ASHRAE. 2004. *Standard 55-2004 - Thermal Environmental Conditions for Human Occupancy*. ASHRAE.

Fountain, M.; Brager, G.; and de Dear, R. 1996. Expectations of indoor climate control. *Energy and Buildings* 24(3):179 – 182.

Froehlich, J. E.; Larson, E.; Campbell, T.; Haggerty, C.; Fogarty, J.; and Patel, S. N. 2009. HydroSense: Infrastructure-Mediated Single-Point Sensing of Whole-Home Water Activity. In *Ubicomp '09: Proceedings of the 11th international conference on Ubiquitous computing*, 235–244. New York, NY, USA: ACM.

Jiang, X.; Dawson-Haggerty, S.; Dutta, P.; and Culler, D. 2009. Design and Implementation of a High-Fidelity AC Metering Network. In *IPSN '09: Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*, 253–264. Washington, DC, USA: IEEE Computer Society.

Kim, Y.; Charbiwala, Z. M.; Singhania, A.; Schmid, T.; and Srivastava, M. B. 2008. SPOTLIGHT: Personal Natural Resource Consumption Profiler. In *HotEmNets 2008*.

Kim, Y.; Schmid, T.; Charbiwala, Z. M.; and Srivastava, M. B. 2009. ViridiScope: Design and Implementation of a Fine Grained Power Monitoring System for Homes. In *Ubicomp '09: Proceedings of the 11th international conference on Ubiquitous computing*, 245–254. New York, NY, USA: ACM.

Poupart, P. 2005. *Exploiting structure to efficiently solve large scale partially observable markov decision processes*. Ph.D. Dissertation, Toronto, Ont., Canada, Canada.

Poupart, P. 2010. Symbolic Perseus. `http://www.cs.uwaterloo.ca/~ppoupart/software.html/`.

Smith, T., and Simmons, R. G. 2004. Heuristic search value iteration for POMDPs. In *Proc. Int. Conf. on Uncertainty in Artificial Intelligence (UAI)*.

Smith, T., and Simmons, R. G. 2005. Point-based POMDP algorithms: Improved analysis and implementation. In *Proc. Int. Conf. on Uncertainty in Artificial Intelligence (UAI)*.

Smith, T., and Simmons, R. G. 2006. Focused real-time dynamic programming for MDPs: Squeezing more out of a heuristic. In *Proc. Nat. Conf. on Artificial Intelligence (AAAI)*.

Smith, T. 2010. ZMDP Software for POMDP and MDP Planning. `http://www.cs.cmu.edu/~trey/zmdp/`.

Spaan, M. T. J., and Vlassis, N. 2005. Perseus: Randomized point-based value iteration for POMDPs. *Journal of Artificial Intelligence Research* 24:195–220.

UNEP. 2007. *Building and Climate Change: Status, Challenges and Opportunities*. United Nations Environment Program.