Model-free Bandit Algorithms for Efficient Temperature Regulation in Buildings

Hadeed Khan

Received August 15, 2024
Accepted October 27, 2024
Electronic access November 30, 2024

In recent years, areas like Texas have seen energy demands skyrocket, often outstripping the available supply due to population growth and extreme weather, and it is not just areas like Texas. Energy consumption is at an all-time high, increasing the usage of nonrenewable energy sources. This research explores the development of an adaptive energy management system for heating and cooling in residential and commercial structures using machine learning, particularly multi-armed bandits (MAB). Our study aims to optimize energy consumption while maintaining desired indoor temperatures in stochastic environments. Through simulations and analysis, we discover that simple statistical approaches with few assumptions tend to outperform more complex ones, emphasizing the importance of balancing exploring new strategies and exploiting existing successful methods. The implications of this research suggest a global energy reduction of up to 3.75%, increasing sustainability and minimizing costs in buildings, particularly in regions with high energy demand.

Keywords: Robotics and Intelligent Machines; Machine Learning; Machine Learning in Energy Management; Multi-Armed Bandits; Using Multi-Armed Bandits in Smart HEMS (Home Energy Management Systems)

Introduction

Smart thermostats are important components in energy management for residential and commercial structures. Many researchers have already explored statistical methods and various aspects of optimization for reducing energy usage in homes. Recent research by Zhou et al. discusses using smart home energy management systems (HEMS). They studied HEMS' intelligent use of electricity and demand response, along with its functionality with items from smart thermostats to home appliances. They found a fraction of American households already use smart thermostats.1 Based on their analysis, this fraction is only increasing. Fueling this growth is the finding that smart thermostats reduce energy usage by about 10-15% using factors such as humidity, hourly energy rates, and temperatures of specific rooms[1]. Similarly, Saha et al. consider additional aspects of smart HEMS like charging, electric cars, and powering appliances rather than just heating and cooling. This finding considers energy usage, discovering that heating and cooling systems use more energy than any other appliance[2].

Additionally, scientist Sarah Royston analyzes the efficiency of heat flows in a home by looking at thermal management and how "experience-based know-how is shaped by changing social, material, and geographical contexts"[3]. Now, if we look at all three of these research papers, we see that they all have a common goal – to reduce energy usage in homes. We currently rely heavily on fossil fuels of which there is a limited supply. Energy usage in American and European homes is at a peak. On top of that, evidence gathered by Saha et al. suggests our carbon footprint is only increasing. Namely, the increasing adoption of electric vehicles, new electrical household appliances, and an exponentially growing population all yield higher energy demand[2]. The heating and cooling system is one of the most energy-consuming in most houses[2]. If we can develop software that distributes heat more efficiently with a smart thermostat, it would not only drastically reduce our carbon footprint but also reduce energy costs and get us to desired temperatures faster. While smart thermostats have taken the initial steps toward intelligent energy management, our program represents a revolutionary advancement in this field. To the best of our knowledge, no other project has implemented a multi-armed bandits-based approach for adaptive energy management in such a comprehensive and scalable manner.

Implicit in our examination of smart thermostats is that heat-flow models can significantly improve energy consumption. To quantify how much energy can be saved, Norford et al. found that "almost 25% of the energy produced worldwide is used to heat and cool homes and commercial buildings"[4]. If our proposed systems can initially save up to 15% on energy costs, our simulation could save up to around 3.75% of worldwide produced energy if our program is implemented in every thermostat and heating/cooling system. We acknowledge this reality is intractable but mention this deduction to define an upper bound for our energy contributions.

This research aims to develop an adaptive energy management system for heating and cooling. By leveraging logic from multi-armed bandits, we explore whether an adaptive energy management system effectively optimizes temperature control. In particular, this system leverages multi-armed bandits for heating and cooling spaces of semi-arbitrary size in stochastic environments that attempt to mimic real-world conditions. The rising demand for energy caused by a growing population stresses energy grids and current non-renewable reserves.

This technology is especially important in regions such as Texas, which receives more demand than supply on the energy grid when temperatures reach extremes. Other areas with extreme temperatures could greatly benefit from our analysis, though our solutions can be adapted to all locations requiring heating and cooling. We define and evaluate novel bandit-based energy management systems that learn over time by modeling the explore-exploit trade-off defined below. We conduct our evaluations with simulations in Python that utilize and extend upon existing multi-armed bandit algorithms. To ensure the practicality of our simulations, we set up the layout of the simulated space as a series of connected rooms with doors and heating sources.

We admit several simplifying assumptions to this setup but maintain our work suitably reflects real-world scenarios in a way that allows our results to generalize. We intend to ensure our software can be tractably implemented in programmable thermostats to reduce energy usage. We accomplish this by using our simulation framework to find the optimal heating and cooling pattern to get us to the desired temperature. This optimization process relies on the principle of multi-armed bandits, which involves exploring different strategies with exploiting the most promising ones. By learning from past experiences, our system adapts to varying environmental conditions to minimize energy usage while maintaining comfort levels defined by user-specified temperatures. In the following sections, we delve deeper into the methodologies employed and the results obtained from our energy management research.

Materials and Methods

A Model-free Approach to Home Energy Management Systems

There are two main types into which commonly utilized methods for HEMS can be grouped. One models everything about the environment and constructs a simulation that attempts to analytically (or semi-analytically) compute optimal energy transfer as a function of the law of physics. However, in real-world cases, this is an incomplete and expensive approach. Many factors, such as insulation in the walls and imperfections in residential structures, must be accounted for. Instead, we prefer to assume nothing about the environment (model-free) and learn primarily through trial and error, a process modeled well by the multi-armed bandit literature. This way, we don't have to account for additional quantities that could affect our system's outcome. Instead, we assume only knowledge about the temperature in each room at a specific time, along with one temperature gauge for outside conditions, making it easy to determine where heating and cooling must be applied as shown in Figure 1. This approach makes our program easier to apply to homes everywhere with just a smart programmable thermostat without requiring an unnecessary number of sensors to be installed throughout the user's homes. Additionally, in many cases, model-free simulations are computationally faster to conduct than the analytical equivalents, accelerating research.

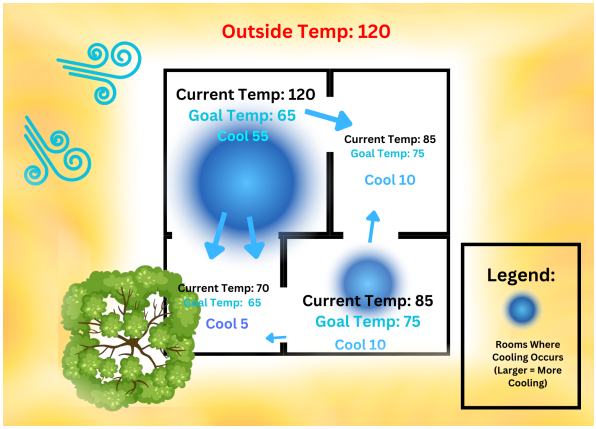

Fig. 1 Diagram of efficient cooling of a 4-room building. Each room has a different current and goal temperature that the user sets. Factors like outside temperature, wind, and tree shade play factors in cooling and the home cooling system. Sending more cool air to certain rooms over others and having cool airflow through the building result in efficient cooling. Measurements shown are in Fahrenheit.

Theoretical Methodology and Problem Setup

In this simulation, we model the energy management problem as a stochastic optimization task, where we have to minimize our energy consumption while maintaining desired indoor temperatures. Let's say $T(r_i,t)$ is the temperature in the room r_i at time t, a random variable following a probability distribution $P(T_{(r_i)},t|T_{(r_{(-i)})})$, where $r_{(}-i)$ represents the set of all room temperatures except the one of interest. Let's also define the energy applied to a room at time $E_{t_i} = E(r_i,t)$, which can imply heating or cooling. Taken together, $R(,T_{(r_1)},...)$ represents the reward function of the temperature of all rooms given the actions taken action taken at the time t, which implicitly depends on previous actions and their results. We aim to find optimal methods of dispensing energy to achieve desired building temperatures while consuming the least amount of energy possible. We approach this task of decision-making under uncertainty with multi-arm

bandits. It is important to note that for our simulations, we relied on the central limit theorem, which suggests that the sum of many small, independent factors affecting temperature can be approximately modeled as a normal distribution. The parameters of the distribution, such as the mean and variance, were estimated using historical temperature data from similar environments, ensuring that the model accurately reflects typical temperature variations in the rooms.

Multi-armed bandit (MAB) algorithms provide a framework for sequential decision-making under uncertainty that we can quickly assess through simulation. At each step, our system has to choose an action from a set of options (also known as the "arms" of the MAB), to choose the arms to give us the highest possible return, an accumulation of reward. We can either stay consistent with one choice and get an expected reward, or we could take a risk and explore a new arm which could be better or worse. A colloquial example of this is eating at restaurants. Let's say every Friday afternoon you eat out. You could eat at your favorite restaurant like usual (exploit), where you know the food is pretty good, or try a new restaurant (explore), which could be much better and become your new favorite, or it could be much worse. Our systems and MAB algorithms repeatedly make explore v. exploit decisions and are evaluated over a long period. This predisposes MAB algorithms to be very useful for heating and cooling requirements.

Many MAB strategies could be useful for model-free energy management with smart thermostats. We define a simple model that applies a fixed amount of energy. We call this the Constant Model, as it represents a simple baseline where the same control settings are applied consistently without exploration. On the other hand, we can also utilize the ε-Greedy Model. This model mainly exploits but occasionally explores an epsilon percentage of the time, when it selects a random action, allowing the system to try and find an arm with a better reward. This model also has variants like the decayed ε-Greedy Model, which adaptively adjusts the exploration rate over time to balance the trade-off between exploration and exploitation as the system learns. Another popular MAB strategy is the UCB or Upper Confidence Bound method. This method balances the exploration of uncertain actions with the exploitation of actions that appear to have higher expected rewards using a specific modifiable formula depending on the variation of UCB. UCB selects actions based on their estimated values and the uncertainty associated with those estimates, prioritizing actions with higher uncertainty and upper-bound reward potential for exploration. Over time, UCB refines its estimates and becomes more accurate in picking the best possible action. By using such MAB algorithms in our system tailored for energy management, we aim to have a balance between exploration and exploitation, ultimately maximizing efficiency while maintaining the desired temperatures.

Mohri et al. look at different MAB strategies and how they work in their paper. He tests different strategies, like ε-Greedy and SoftMax. The ε-Greedy strategy primarily exploits known successful actions but occasionally explores new ones with a small probability (epsilon), allowing for a balance between exploration and exploitation. On the other hand, the SoftMax approach assigns probabilities to each action based on their expected rewards, selecting actions with higher probabilities while still allowing for the exploration of less promising options. He also introduces a new one called Poker. He finds that simpler strategies often do better than the more complicated algorithms[5]. For example, a finding in Mohri's data is that the ε-greedy Model, a much simpler model than the others, far outperformed other models like the Interval Estimation model, a complicated algorithmic intensive model[5]. The estimation model assigns each lever to an optimistic reward estimate within a confidence interval. It prioritizes the selection of levers with higher optimism to balance out exploration and exploitation in these multi-armed bandit problems. The experiments show that methods like Interval Estimation and Poker shine, especially with dynamic data[5]. For our specific topic of energy management, papers by Kaza et al. provide valuable insights into multi-armed bandits for energy management. In exploring bandit algorithms, this research implies that a promising approach for our model-free simulation experiments would be to use bandit algorithms like these as they do not need as many sensors and factors to consider while still achieving an accurate and efficient end result[6].

Implementation

Our program is designed to be implemented in houses, apartments, offices, or other buildings with different structures using a smart programmable thermostat and a small number of temperature sensors, one per room. Over 16% of American households already have this, implying that adopting novel smart thermostat technology could be fast. Importantly, there is no additional cost for this program as long as they already have a smart programmable thermostat with the required sensors and a Wi-Fi connection.

We designed a simulation framework, FlexSim, in Python that allows for visualization and easy changes in the home layout to better simulate the most efficient heating/cooling arrangement. FlexSim offers ease of modification and incorporates key parameters such as room dimensions, the number of rooms, building shape, size, external temperatures, and initial indoor temperatures for each room. We assumed constant occupancy to simplify the modeling process. Each MAB algorithm was specifically created for our study, with separate Python files to enhance modularity and reference ease. This framework enabled us to test a wide range of building layouts, including offices, hotels, small houses, large mansions, and stores, allowing us to evaluate how the energy management system adapts effectively to different environments.

Our research incorporates theoretical frameworks from multi-armed bandit literature to model and analyze the stochastic nature of heating and cooling systems. Leveraging these mathematical concepts allows us to comprehensively understand the dynamics involved in adaptive energy management. Using these theories, we aim to enhance the efficiency and effectiveness of energy consumption in residential and commercial heating, contributing to sustainable and optimized energy usage. We also include several intuitive but naive baselines on which we compare the results of our candidate methods.

We will employ software languages and tools such as Python and NumPy, a scientific computing library, to implement our algorithms for the simulations. Python's versatility and NumPy's numerical capabilities make them ideal for conducting simulations and analyzing large datasets, providing a robust foundation for our research. Utilizing these tools helps with the accuracy and efficiency of our simulations, allowing us to explore a wide range of scenarios and parameters in developing and evaluating our adaptive energy management system. We also have the program save the simulation output in the compressed files we can refer back to, plotting the saved data on charts and graphs and comparing them.

To evaluate and assess the findings of our work we will utilize the reward function of the simulation, which we define to essentially be how close we can get to achieving our desired temperatures and how quickly we do so. This way, we have a quantitative measure we can analyze across structures. We will measure this by looking at how close the actual temperature in each room was to the desired temperature at each time step. The smaller the difference, the higher the reward. We also considered how quickly the system reached the desired temperature, giving lower scores for slower responses. The overall reward was averaged across different room layouts and sizes to account for variations.

Experimental Results

After running our simulation, we can see that the Constant Model was overall the most efficient option with a reward score of 8.8, with the ε-Greedy Model following with a reward score of -0.05, and the UCB Model in last place with a reward score of -1.2. Upon further investigation and breakdown, we noticed that if we separate the results based on whether the model was heating or cooling, we get a better picture of the situation. The Constant Model performed exceptionally with a reward score of 13.9 for heating while earning a poor -5.7 for cooling. The UCB Model, on the other hand, performed a -2.49 for heating with a healthy reward score of 4.5 for cooling. Lastly, despite being the most complex model, the ε-Greedy Model came out with poor reward scores across the board, with a -0.13 in heating and a -0.43 in cooling.

There are multiple insights we can take away from these experimental results. First, simpler models like the Constant Model are more efficient than more complicated models such as the ε-Greedy Model. Intuitively, across the board, we saw that the bigger the room, the more energy it takes to heat or cool. There were a few exceptions when neighboring rooms were at a convenient temperature and assisted with regulation. Additionally, layouts with the right balance of clustered and spanned out achieved the best efficiency with heating and cooling. All models were tested under many realistic conditions, including scenarios with rapid external temperature changes. Our results reflect the model's performance across all these varied conditions. Our findings indicate that the model maintains its efficiency even when faced with natural, rapid temperature changes and environmental variability.

In the end, enabling a combination of the UCB model for cooling and the Constant Model for heating would have given us a total average reward score of about 11.56, making for an efficient combination.

Discussion

After running our simulation, we can see that the Constant Model was overall the most efficient option with a reward score of 8.8, with the ε-Greedy Model following with a reward score of -0.05, and the UCB Model in last place with a reward score of -1.2. Upon further investigation and breakdown, we noticed that if we separate the results based on whether the model was heating or cooling, we get a better picture of the situation. The Constant Model performed exceptionally with a reward score of 13.9 for heating while earning a poor -5.7 for cooling. The UCB Model, on the other hand, performed a -2.49 for heating with a healthy reward score of 4.5 for cooling. Lastly, despite being the most complex model, the ε-Greedy Model came out with poor reward scores across the board, with a -0.13 in heating and a -0.43 in cooling.

There are multiple insights we can take away from these experimental results. First, simpler models like the Constant Model are more efficient than more complicated models such as the ε-Greedy Model. Intuitively, across the board, we saw that the bigger the room, the more energy it takes to heat or cool. There were a few exceptions when neighboring rooms were at a convenient temperature and assisted with regulation. Additionally, layouts with the right balance of clustered and spanned out achieved the best efficiency with heating and cooling. All models were tested under many realistic conditions, including scenarios with rapid external temperature changes. Our results reflect the model's performance across all these varied conditions. Our findings indicate that the model maintains its efficiency even when faced with natural, rapid temperature changes and environmental variability.

In the end, enabling a combination of the UCB model for cooling and the Constant Model for heating would have given

us a total average reward score of about 11.56, making for an efficient combination.

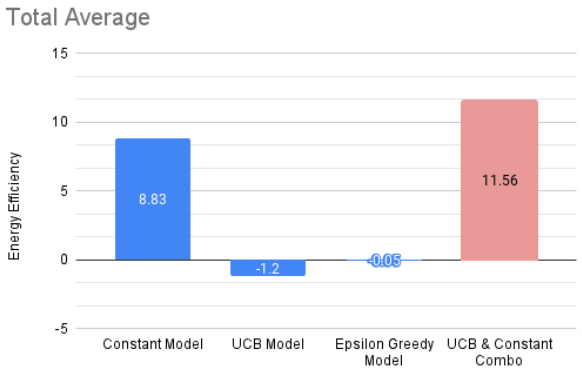

Fig. 2 Demonstration of overall Energy Efficiency from the 3 models tested on multiple building layouts, resulting in a reward score. The use of UCB for cooling and the Constant Model for Heating as one combination model is also shown and outperforms the other individual models.

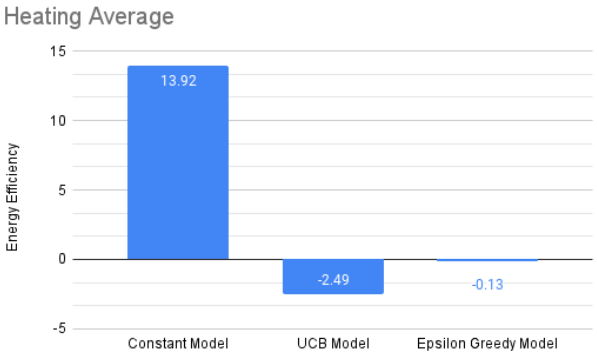

Fig. 3 Demonstration of Heating Energy Efficiency from the 3 models tested on multiple building layouts such as different homes, apartments, offices, and hotels, resulting in a reward score. The Constant Model outperforms UCB and the ε-Greedy Model and is the only model that receives a positive reward score.

Conclusion

In conclusion, our study aimed to develop an adaptive energy management system for heating and cooling by leveraging concepts from multi-armed bandits. Through simulations and analysis, we found that simpler models, particularly the Constant Model, outperformed more complex ones in optimizing residential and commercial heating and cooling energy consumption. Our findings also emphasize the importance of using separate models for heating and cooling to achieve maximum efficiency, as seen in Figure 5. The implications of our research extend

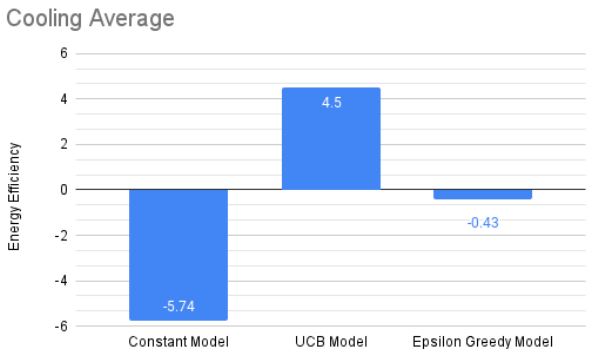

Fig. 4 They demonstrated Cooling Energy Efficiency from the 3 models tested on multiple building layouts such as different homes, apartments, offices, and hotels, resulting in a reward score for each model. The UCB Model outperforms the Constant Model and ε-Greedy models and is the only model that receives a positive reward score.

to the broader goal of reducing energy usage and increasing sustainability in homes, offices, and other buildings, especially in regions with high energy demand. By incorporating adaptive energy management strategies, households can potentially reduce their carbon footprint and energy costs while maintaining comfortable indoor temperatures.

Acknowledgments

I would like to thank my research mentor, Samuel Showalter, for supporting me in my understanding of machine learning techniques and the execution of a Multi-Armed-Bandit Model. I would also like to thank Tyler Moulton for helping me through the submission process whenever I needed help. Lastly, I would like to thank my parents for their support and motivation in making this paper possible.

Authors

Hadeed Khan is a senior at the Liberal Arts and Science Academy in Austin, Texas. Passionate about Artificial Intelligence and Machine Learning, he plans on majoring in Software Engineering or Computer Science next year.

References

1 B. Zhou, W. Li, K. Chan, Y. Cao, Y. Kuang, X. Liu and X. Wang, *Smart home energy management systems: Concept, configurations, and scheduling strategies.*

2 A. Saha, M. Kuzlu and M. Pipattanasomporn, *Demonstration of a home energy management system with smart thermostat control.*

3 S. Royston, *Dragon-breath and snow-melt: Know-how, experience and heat flows in the home.*

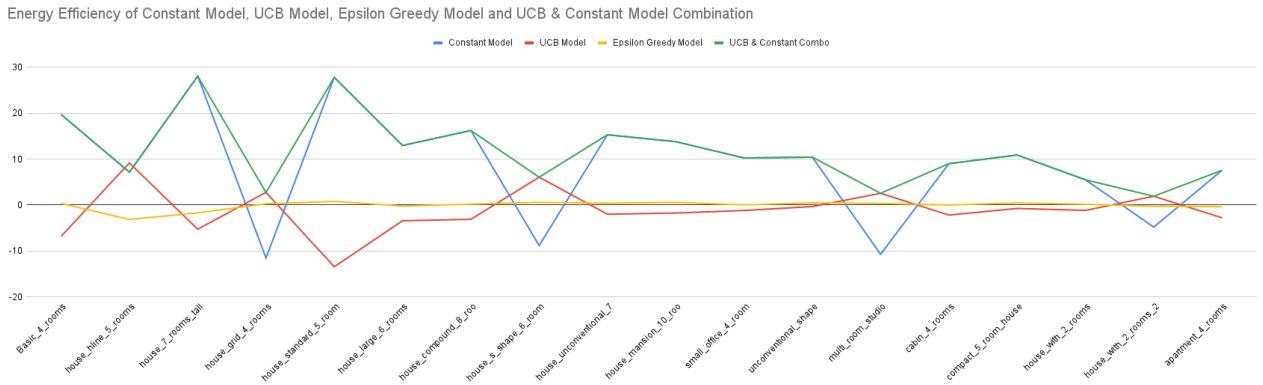

Fig. 5 Demonstration of the Energy Efficiency of the UCB Model, ε-Greedy Model, and Constant Model as well as the combination of the UCB (cooling) and Constant (heating) models shown on the different building simulations with different scenarios like different houses, apartments, offices, hotels and other buildings. The UCB and Constant Model Combination is consistently the most efficient model for the different buildings, while the ε-Greedy Model is consistently close to zero reward score.

4 L. Norford and Gribkoff, *Heating and Cooling*, MIT Climate.

5 J. Vermorel and M. Mohri, *Multi-armed Bandit Algorithms and Empirical Evaluation*, https://doi.org/10.1007/11564096_42., Retrieved from.

6 R. Meshram and K. Kaza, *Simulation Based Algorithms for Markov Decision Processes and Multi-Action Restless Bandits*.