Research on HVAC Occupancy Detection with ML and DL Methods

Dongyang Zhang

College of Civil Engineering, Hunan University， Changsha, Hunan Province, 410002, China

dongyang@hnu.edu.cn

Abstract. It is necessary to optimize Heating, Ventilating and Air Conditioning (HVAC) system efficiency, significantly reduce energy consumption and costs, and minimize carbon emissions by accurately predicting occupancy patterns using advanced Machine Learning (ML) and Deep Learning (DL) techniques. This research focuses on improving HVAC system efficiency through occupancy detection using ML and DL techniques. The study addresses the critical issue of high energy consumption in buildings, which accounts for about 40% of total energy use, by optimizing HVAC operations to reduce waste and carbon emissions. By predicting occupancy patterns accurately, HVAC systems can be adjusted to provide heating and cooling only when necessary, leading to significant energy savings and cost reductions. The research employs various predictive models, including regression, time series forecasting, and ensemble methods, achieving high accuracy rates, particularly with K-Nearest Neighbors (KNN). Despite their complexities and challenges, advanced control methods like Model Predictive Control (MPC) and Reinforcement Learning (RL) are also explored. Overall, the study highlights the potential of integrating advanced data analysis and predictive modeling to enhance building energy management, promoting more sustainable and environmentally friendly practices.

Keywords: HVAC, Machine Learning, Deep Learning, Green Building

1. Introduction

The control and regulation of the human living environment in buildings are essential for ensuring the comfort and well-being of occupants. However, adjusting the temperature and humidity within buildings often consumes a significant amount of energy. In recent years, the intensifying phenomenon of global warming and increasingly abnormal climate changes have underscored the urgency of addressing environmental protection and resource conservation. It is necessary to conduct a comprehensive assessment of traditional Heating, Ventilating and Air Conditioning (HVAC) engineering in buildings.

The building sector is one of the largest consumers of energy, accounting for approximately 40% of total energy use worldwide. HVAC systems are a significant component of this consumption, as they are responsible for regulating the indoor climate to meet the thermal comfort needs of occupants. The increasing awareness of

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_69

the environmental impact of energy consumption, coupled with the economic benefits of reducing operational costs, has driven the need for more efficient HVAC systems. Traditional HVAC systems often operate based on preset schedules or simple feedback mechanisms that do not account for actual occupancy patterns or varying environmental conditions, leading to inefficient energy use [1].

Numerous studies have explored various approaches to improving the energy efficiency of HVAC systems. Recent advancements in machine learning (ML) and deep learning (DL) have opened new avenues for enhancing HVAC operations. These technologies enable the development of predictive models that can forecast occupancy patterns and adjust HVAC settings in real-time, thereby optimizing energy use. Additionally, advanced control strategies such as Model Predictive Control (MPC) and Reinforcement Learning (RL) have been investigated for their potential to manage the complex dynamics of HVAC systems more effectively. Despite these advancements, challenges remain in achieving reliable and scalable solutions that can be widely adopted in existing buildings. This assessment aims to summarize the general energy consumption patterns of HVAC systems and to utilize advanced methodologies such as deep learning and reinforcement learning to optimize the use and distribution of energy in buildings. The ultimate goal is to enhance energy efficiency and reduce carbon emissions while maintaining a comfortable living environment.

This research focuses on leveraging ML and DL techniques to improve the efficiency of HVAC systems through accurate occupancy detection and real-time adjustments, aims to provide a comprehensive evaluation of traditional HVAC engineering, identify common energy consumption patterns, and propose optimization strategies using advanced ML and DL methods. The ultimate objective is to achieve higher energy efficiency and reduced environmental impact in building operations.

2. Data and Methods

2.1 Data

In this research, a dataset which includes figures for building energy cost performance was used to predict building's loads in heating and cooling. There are 4 types of building cooling strategy in the dataset: Mechanically ventilated, Naturally ventilated, Air conditioned and Mixed model [2,3]. This dataset is characterized by its detailed temporal and categorical data, allowing for a thorough analysis of energy usage patterns.

Table 1. Example of air-conditioning load relative with house size

Relative Compactness	Surface Area	Wall Area	Roof Area	Overall Height	Orientation	Heating Load	Cooling Load
0.7638	514.3	294.0	110.25	7.0	2	15.55	21.33
0.9800	514.3	294.0	110.25	7.0	3	15.55	21.33
0.9800	514.3	294.0	110.25	7.0	4	15.55	21.33
0.9800	514.3	294.0	110.25	7.0	5	15.55	21.33
0.9000	563.7	318.5	122.50	7.0	2	20.84	28.28

Table 1 shows an example of how the size of a house can affect its air-conditioning load which calculated by kW(kilowatt) terminally. Figures in the table show that the load would increase with the size growing.

2.2 Data Pre-processing

Data preprocessing involved cleaning the dataset to remove inconsistencies and handling missing values to ensure data integrity. Normalization techniques were applied to scale the data, facilitating better performance of machine learning algorithms.

All text files have seven columns as date, temperature (in Celsius), humidity (in percentage), light (in lux), carbon dioxide (in ppm), humidity ratio (in kg vapor per kg air) and occupancy (1 or 0 for occupied or not occupied).

For training and testing the models, 8143 instances will be used as training, dataset (2665 instances) as validation and datatest2(9752 instances) as test data.

There are plenty of missing values in the dataset. The first step of my data analysis process is to see whether it's conventional to replace the missing values with the mean.

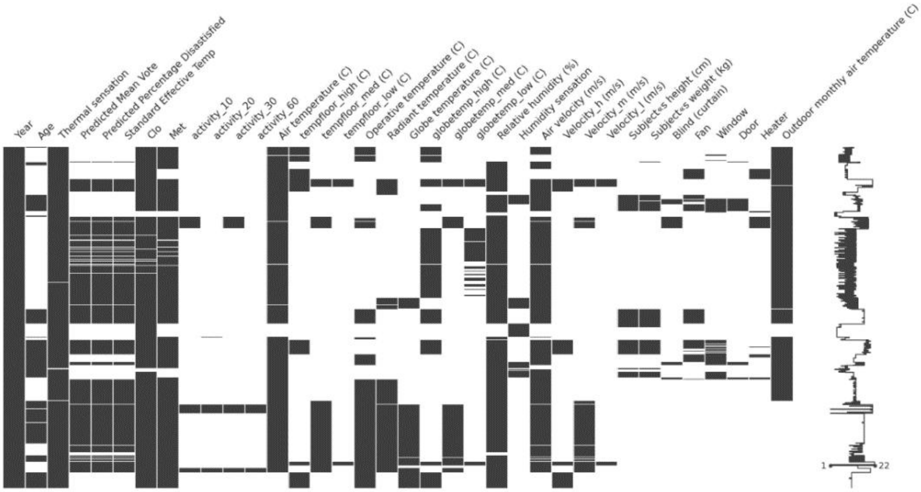

Fig.1. Some defaults in the dataset(Photo/Picture credit : Original)

As showed in Fig. 1, there are a large number of discrete and useless data items in the dataset, and we need to isolate the useful data items before feeding the data into the model.

As the results of the analysis show, the correlations between occupancy light value is more correlated with occupancy than others.

2.3 Model

Broad sense buildings account for approximately 40% of total energy consumption. Predicting the heating and cooling loads during the initial design phase is crucial for identifying optimal solutions among various designs, as well as for ensuring efficient energy use during the operational phase post-construction. This research applied several models, including linear regression, decision trees, and support vector machines (SVM), to predict the heating and cooling loads based on a dataset for building energy performance.

These algorithms were chosen for their ability to handle large datasets and provide accurate predictions. Linear regression is useful for understanding relationships between variables, decision trees offer easy interpretation and handle non-linear relationships, and SVM is effective in high-dimensional spaces. The models aimed to predict thermal energy measures — cooling load (CL) and heating load (HL) — regulated by HVAC systems to maintain desirable indoor conditions.

Model Predictive Control, MPC. The operational characteristics of HVAC systems are highly complex. Most related research employs an optimization control method known as MPC, which relies on mathematical optimization to address control issues in air conditioning systems. This method necessitates a "low-order" mathematical model, typically linear, to represent the system. However, a complete commercial central air-conditioning system encompasses numerous different devices and operates under constantly changing external conditions, making it challenging to capture such complexity with a "low-order" model. Although various workarounds have been proposed in the literature to address this issue, they often prove too cumbersome for practical application [4].

Reinforcement Learning, RL. RL is a machine learning technique that enables autonomous agents to learn optimal behavior through interaction with an environment, guided by rewards and penalties. In applications such as air conditioning system control, RL offers the ability to dynamically adjust settings to meet specific goals like energy efficiency or user comfort. But, RL's reliance on trial-and-error exploration poses challenges in real-world scenarios where mistakes can be costly, such as potential damage to equipment or discomfort to occupants. To mitigate these risks, a common approach involves initially training RL agents in simulated environments where learning can occur safely and efficiently. Once trained, the agent can then transfer its learned policies to real-world settings, leveraging its experience to make informed decisions while minimizing adverse impacts. This hybrid approach balances the benefits of RL's adaptive learning capabilities with the practical constraints of operational environments, ensuring effective and reliable control strategies. By harnessing RL, air conditioning systems can achieve optimal performance over time, adjusting to varying conditions and preferences without constant human intervention, thereby enhancing efficiency, reducing operational costs, and improving overall system reliability in diverse applications ranging from commercial buildings to industrial facilities [5].

Reinforcement learning (RL) offers the significant advantage of learning without requiring explicit models. However, this strength also introduces several disadvantages, particularly in context of HVAC system control. The primary challenge is that RL's learning process necessitates a comprehensive exploration of the environment, inherently involving trial and error. This approach means that mistakes are an integral part of the learning process, as identifying optimal solutions is impossible without the ability to make and learn from errors.

For large, centralized HVAC systems, the cost of such mistakes is prohibitively high. Many components within these systems have specific operating conditions that, if exceeded, can cause significant damage. For example, a large chiller requires a precise amount of cooling and specific start up times. Frequent starting and stopping of the chiller can make its motor overheating, causing potentially irreparable damage [6]. Furthermore, commercial buildings must meet stringent air conditioning demands. In office buildings, overcooling or overheating can lead to tenant dissatisfaction, while in industrial buildings, unmet cooling requirements can damage precision equipment.

Given these constraints, conducting full-scale online intensive learning is impractical and risky. Convincing building owners to adopt such a solution is challenging due to the potential risks and costs involved. Therefore, an optimal solution involves training an RL agent in a simulated environment before deploying it in a real-world setting. This approach allows the agent to learn and refine its control strategies without the risk of causing damage or inefficiency in actual operations.

Despite the use of a simulator, it is important to note that RL remains fundamentally a "model-free" algorithm. The complexity of the simulator does not impact the effectiveness of the training process, as RL does not rely on pre-defined models but instead learns directly from interactions within the simulated environment. The method mentioned above ensures that once the training is complete, the RL agent can implement effective control strategies in real-world HVAC systems, balancing the need for efficient operation with the minimization of risks and costs.

2.4 Evaluation

Evaluation metrics included Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2). MAE measures the average magnitude of errors in predictions, RMSE provides the standard deviation of prediction errors, and R^2 indicates how well the data fits the model. Assessing the accuracy and reliability of predictive models is crucial for optimizing HVAC system designs and enhancing building energy performance. By employing comprehensive metrics, this approach ensures more sustainable energy consumption and contributes to the development of more efficient HVAC systems [7].

3. Data Processing and Analysis

3.1 Data Pre-processing

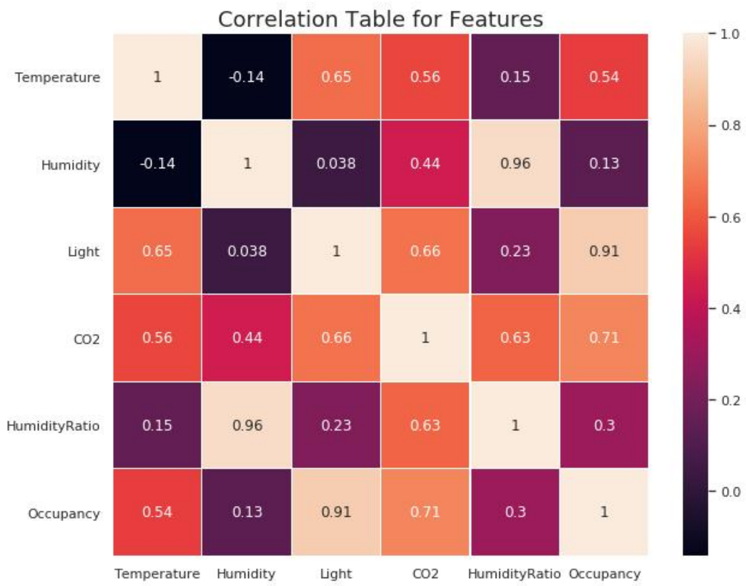

Fig.2 Confusion matrix shows the correlation coefficients between various features in the dataset.(Photo/Picture credit : Original)

Fig. 2 illustrates the correlation coefficients among various features in the dataset, highlighting the strength and direction of the relationships between factors such as temperature, humidity, light, levels of carbon dioxide, humidity ratio, and occupancy.

This correlation matrix illustrates the relationships between various features such as temperature, humidity, light, CO2, humidity ratio, and occupancy, revealing how strongly they are correlated with one another. These insights are crucial for the research as they guide the selection and engineering of features, helping to build more accurate and effective predictive models for optimizing HVAC system performance and improving energy efficiency in buildings.

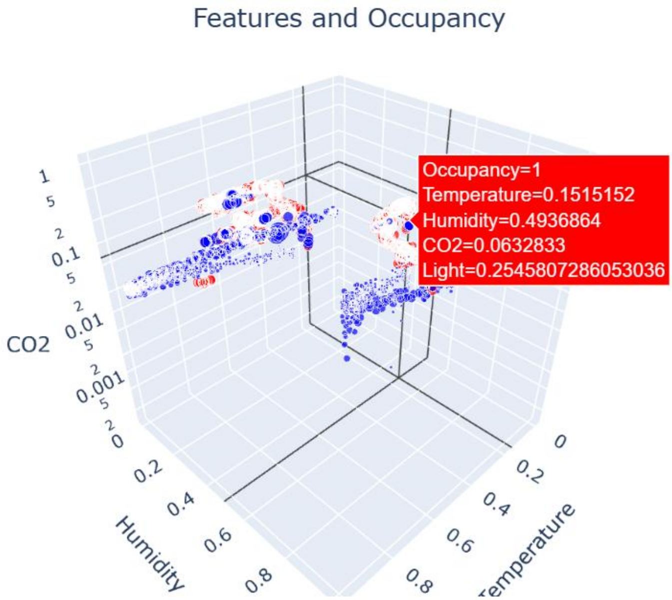

Fig. 3 Data spots show good spatial correlation(Photo/Picture credit : Original)

By using the light value as 4th dimension, Fig.3 shows that the higher light will lead to bigger dots and the lower light will lead to smaller dots.

3.2 Classification with Machine Learning Methods

K-Nearest Neighbors (KNN) Method. In the study, the performance of KNN algorithm was evaluated using different distance metrics and weighting schemes to determine the optimal configuration for predicting HVAC system efficiency (Fig.4). The analysis revealed that when the k-value is low, the Manhattan distance metric outperforms others, providing more accurate predictions. This indicates that for smaller neighborhoods, the Manhattan distance, which calculates the sum of absolute differences between points, is more effective in capturing the relevant patterns and variations in the data. While as the k-value increases, the Euclidean distance metric becomes the better option [8]. This suggests that for larger neighborhoods, the Euclidean distance, which calculates the straight-line distance between points, is more suitable for capturing the overall structure and relationships within the dataset. This transition highlights the importance of selecting the appropriate distance metric based on the size of the neighborhood considered by the KNN algorithm.

Fig.5 presents the analysis of validation data using a KNN model. The rows correspond to the actual classes, while the columns represent the classes predicted by the KNN model. The value 1616 in the top-left corner indicates True Negatives (TN), where the actual class is 0 and the model correctly predicted them as 0. Similarly, the value 952 in the bottom-right corner represents True Positives (TP), where the actual class is 1 and the model correctly predicted them as 1. The values 77 and 20 indicate instances where the actual classes were 1 or 0, respectively, but the model incorrectly predicted them as 0 or 1 [9].

Additionally, the study found that using uniform weights, where each neighbor contributes equally to the prediction, resulted in better performance compared to distance-weighted schemes. This finding suggests that in this particular application, the influence of each neighbor on the prediction should be treated equally, regardless of their distance from the target point. These insights are crucial for optimizing the KNN algorithm, enhancing its predictive accuracy, and ultimately improving the energy efficiency of HVAC systems in buildings [10,11].

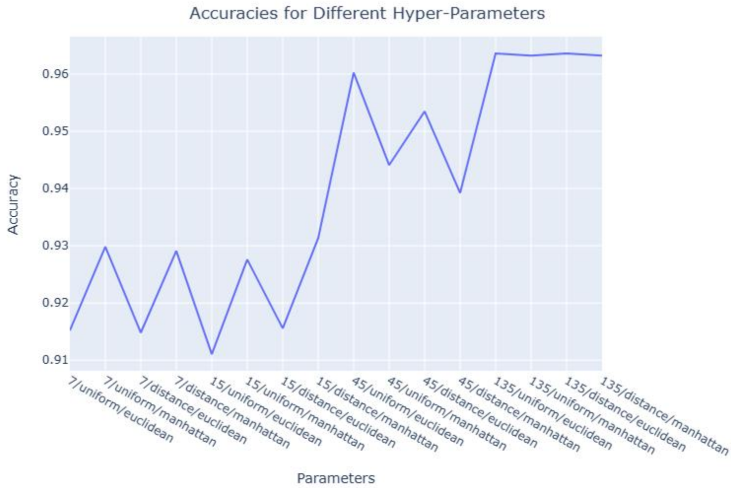

Fig.4 Accuracies change as the k changes, while it shows increasing revelation overall(Photo/Picture credit : Original)

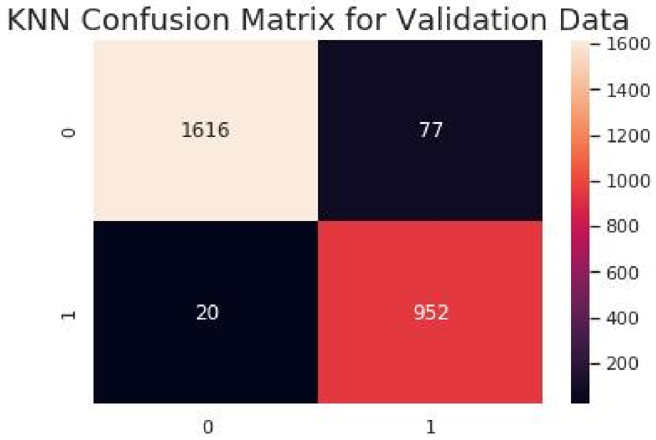

Fig. 5 The confusion matrix shown in the image is a result of analyzing validation data using a K-Nearest Neighbors model.(Photo/Picture credit : Original)

3.3 Shortcoming

The models demonstrated satisfactory predictive capabilities with regard to occupancy. The accuracy of the model is nearly 98%. It's quite hard to figure out whether the suitability of ML or DL for this dataset and problem.

Before figuring out, it would be beneficial to examine the confusion matrix, which is created when evaluating models with the test data. The SVM model appears to exhibit a degree of bias towards the occupied class. It might be presumed that the neural network would be a much more suitable method for achieving more stable and accurate results without significant errors.

4. Conclusion

The practical constraints of existing building infrastructure often preclude extensive modifications due to prohibitive costs and operational disruptions that may not align with management's directives. Therefore, optimizing energy efficiency requires strategies that can work within the constraints of established systems. This research emphasizes the effectiveness of machine learning in analyzing vast datasets collected over extended periods from diverse building types. Such data-driven insights enable models to identify optimal energy usage patterns, thereby fostering adaptive strategies that minimize wastage and enhance overall efficiency.

Moving forward, there are promising avenues for further exploration. For instance, exploring occupant satisfaction with thermal conditions in naturally ventilated buildings compared to mechanically ventilated ones could provide valuable insights into comfort preferences and energy consumption patterns. This line of inquiry acknowledges that naturally ventilated buildings, reliant on external air, may vary significantly in thermal comfort depending on regional climate conditions and architectural design.

By integrating academic rigor with practical considerations, this research advocates for a systematic approach to improving HVAC energy efficiency. It underscores the transformative potential of machine learning when applied to comprehensive datasets gathered over extended periods, illustrating its capacity to optimize operational efficiencies without necessitating disruptive changes to existing infrastructure. This dual focus on empirical evidence and practical applicability positions the study at the forefront of advancing sustainable building practices through innovative technology and rigorous scientific inquiry.

References

1. Hagström, F., Garg, V., & Oliveira, F.. Employing Federated Learning for Training Autonomous HVAC Systems. ArXiv. /abs/2405.00389 (2024)
2. Li, H., & Xu, J.. Priority-based Energy Allocation in Buildings for Distributed Model Predictive Control. ArXiv. /abs/2403.13648 (2024)
3. Bian, Y., Fu, X., Gupta, R. K., & Shi, Y. (2024). Ventilation and Temperature Control for Energy-efficient and Healthy Buildings: A Differentiable PDE Approach. ArXiv. /abs/2403.08996
4. Salter, P., Huang, Q., & Cesar, P.: Quantifying and Predicting Residential Building Flexibility Using Machine Learning Methods. ArXiv. /abs/2403.01669 (2024)
5. Manjavacas, A.: An experimental evaluation of Deep Reinforcement Learning algorithms for HVAC control. ArXiv. /abs/2401.05737 (2024)
6. Clemente, A. V., Nocente, A., & Ruocco, M.: Global Transformer Architecture for Indoor Room Temperature Forecasting. ArXiv. /abs/2310.20476 (2023)
7. Tekler, Z. D., Lei, Y., Dai, X., & Chong, A.: Enhancing personalised thermal comfort models with Active Learning for improved HVAC controls. ArXiv. /abs/2309.09073 (2023)
8. Wang, M., Willes, J., Jiralerspong, T., & Moezzi, M.: A Comparison of Classical and Deep Reinforcement Learning Methods for HVAC Control. ArXiv. /abs/2308.05711 (2023)
9. Liu, R., Pan, Y., & Chen, Y.: Laxity-Aware Scalable Reinforcement Learning for HVAC Control. ArXiv. /abs/2306.16619 (2023)
10. Wang, H., Chen, X., Vital, N., Duffy, E., & Razi, A.: Energy Optimization for HVAC Systems in Multi-VAV Open Offices: A Deep Reinforcement Learning Approach. ArXiv. /abs/2306.13333 (2023)
11. Xiao, S.: Application of energy-saving technologies in HVAC systems for a large complex. Shanghai Gas, 2024(02), 24-28 (2024)