OPTIMIZATION OF OCCUPANT NUMBER PREDICTION MODEL USING HYPERPARAMETER TUNING

Young Jae Choi
First Author, School of Architecture and Building Science, Chung-Ang University, Republic of Korea

Eun Ji Choi & Nam Hyeon Kim & Jae Yoon Byun & Kang Woo Bae
Coauthor, School of Architecture and Building Science, Chung-Ang University, Republic of Korea

Jin Woo Moon
Corresponding author (e-mail: gilerbert73@cau.ac.kr), School of Architecture and Building Science, Chung-Ang University, Republic of Korea,

ABSTRACT: *Recently, occupant-centric control (OCC) has been attracting attention. Occupancy information is essential for OCC; predicting this information enables optimal control of heating, ventilation & air conditioning (HVAC) systems. To this end, attempts have been made to develop an occupant number prediction model using machine learning. However, most studies have focused on performance comparison between models, and the application of hyperparameter optimization has been insufficient. Therefore, in this study, an occupant number prediction model was developed, and an optimal model was derived by applying Bayesian optimization and Hyperband optimization. Through an analysis of the results, the limitations of the occupant number prediction model and the development direction are determined.*

The ground truth data for a medium-sized office building was used for training and the gated recurrent unit (GRU), which is mainly used for time series prediction, was employed; it was trained to predict the average occupant number after 30 minutes. The GRU model was developed in three ways: a base model, Bayesian optimization model, and Hyperband optimization model.

According to the performance evaluation, the base model had the lowest mean absolute error (MAE) of 0.8646 person, and the model with Bayesian optimization presented the lowest root mean squared error (RMSE) of 1.7394 person. Consequently, the model with Bayesian optimization, which has potential for stable predictions, was selected as the optimal model. However, all three models showed high prediction accuracy, and the difference in performance according to optimization was insignificant. Additionally, the shifting phenomenon in which past values were used for the prediction values occurred. These are judged to be due to the uncertainty characteristic of occupancy data. Therefore, appropriate output data should be selected for the purpose of using the occupant number prediction model. In future research, a control algorithm will be developed to remedy the limitations of the prediction model.

KEYWORDS: *Artificial Intelligence, Gated Recurrent Unit, Occupancy Forecasting, Bayesian Optimization, Hyperband*

1. INTRODUCTION

Recently, research on occupant-centric control (OCC) has been pursued as a method for solving problems related to building energy and occupant health. OCC is a control method that provides comfortable indoor environmental quality (IEQ) to occupants and minimizes energy consumption (Park et al., 2019). In order to apply OCC to buildings, the collection of real-time data on the environment, HVAC system, and occupants is essential. The key data among these is the number of occupants. An accurate prediction of this information enables preliminary operation and early shutdown of the system, enabling energy savings through efficient system operation (Choi et al., 2022; Panchabikesan, Haghighat & El Mankibi, 2021).

Probabilistic modeling and data-driven methods have generally been used as a means for predicting the number of occupants. In particular, the machine learning-based data-driven model can adapt to different occupant schedules for each space through learning, and it has been shown to have superior prediction accuracy compared with the probabilistic model. Previous research has shown that artificial neural networks and recurrent neural networks are the most prominent among machine learning algorithms (Hunchuk, Sanner & O'Brien, 2019; Kim et al., 2019; Schiele, Kopema & Brunner, 2021). The prediction performance of neural network models varies depending on the combination of hyperparameters, so various trials and detailed analysis are required during the development process. However, in most previous studies, hyperparameter optimization was not considered, and only superficial performance comparisons of various learning algorithms were performed.

Therefore, the purpose of this study is to derive the optimal model for occupancy prediction by applying various hyperparameter tuning methods during the development stage. For the machine learning algorithm, a gated recurrent unit (GRU), which presents rapid learning and excellent performance in time series prediction, was adopted, and for the hyperparameter tuning method, Bayesian optimization and Hyperband optimization were applied. The results of this study determine the optimal hyperparameter tuning method and provide occupant number prediction through an analysis of the results.

2. METHODOLOGY

2.1 Research method

In this study, the open dataset provided by Luo et al. (2022) was used for the development of the occupant number prediction model. The building being studied is a medium-sized office building with a total floor area of 10,400 m^2 located at the Lawrence Berkeley National Laboratory (LBNL). The dataset provides occupant number data for a specific zone between May 2018 and February 2019. The time step of data is 1-minute intervals. The input data were created by adding information about weekday, hour, and minute to the data, and data pre-processing was performed. The prediction model was developed with GRU, and the model that presented the best performance was derived by comparing the prediction accuracy between the base model, the Bayesian optimization model, and the Hyperband optimization model. A summary of the research method is shown in Fig. 1.

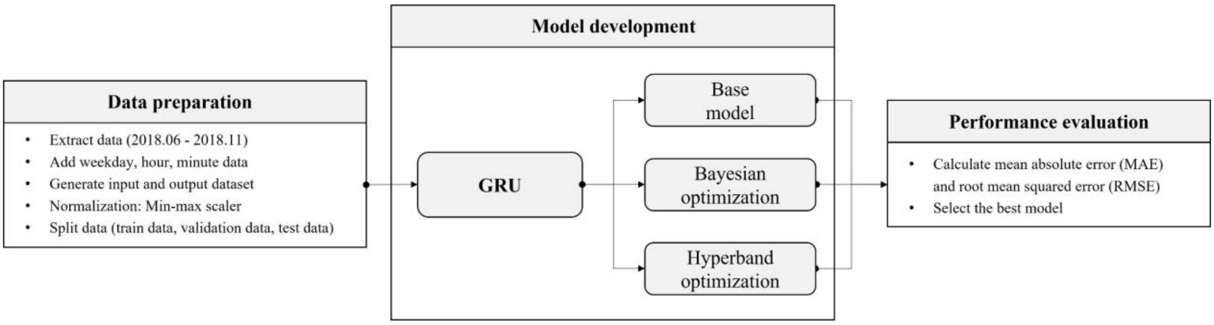


Fig. 1 Research method

2.2 Data preprocessing

The data preprocessing was divided into five steps: data extraction, temporal information addition, input and output data generation, normalization, and data segmentation processes.

First, data extraction was conducted on data for the number of occupants between June 2018 and November 2018. The interval between data points in the extracted data is 1 minute, and the total number of data points is 262,081.

Second, the weekday, hour, and minute data, which are temporal information, were added to the occupant number data. In the case of offices, most show a regular pattern of occupancy for the weekday and daily schedules, and the number of occupants is affected by the office schedule. Therefore, in order to predict the future occupant number, temporal data should be included in the input data.

Third, to construct the input and output data, a prediction target was selected, and a sequence of input data was generated. Because the occupancy prediction model will be applied to predictive HVAC control, the average occupant number of 30 minutes after, which is the following HVAC control cycle, was selected as the output data. The reason for predicting the average value rather than the occupant number after 30 minutes is that the larger the number of occupants, the more important the size of the occupant number and the trend in increase or decrease rather than the exact number of occupants. In the case of input data, data collected at 5-minute intervals were used because the learning time would be longer if historical data at 1-minute intervals were used, and a window size of 12 which is the number of historical data was set to reflect the history of the past hour.

Fourth, because there are no negative numbers in the time and occupant number data, the min-max scaler was applied during data normalization to obtain a value between 0 and 1.

Finally, data preprocessing was completed by dividing the normalized data into 60%, 20%, and 20% portions, respectively, to generate the train, validation, and test datasets. Because the number of occupants represents a continuous value with respect to time, random shuffling was not applied. The preprocessed data structure and prediction target are summarized in Fig. 2.

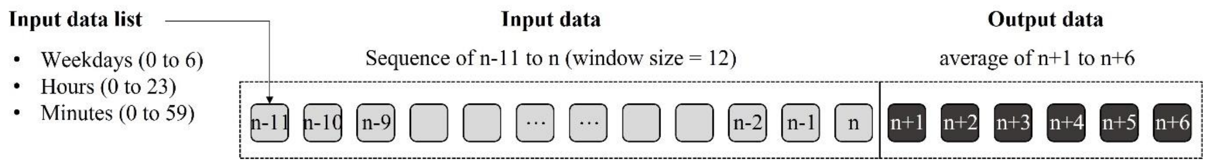


Fig. 2 Structure of input and output data

2.3 Development of occupancy prediction model

This study employed the GRU model, which has been adopted in various time series prediction problems. GRU was proposed by Cho et al. (2014) as a type of recurrent neural network. Relative to long short-term memory (LSTM), which is advantageous for long-term memory, GRU has a simple structure and few parameters, so it is known to have a fast training speed and excellent predictive performance. GRU consists of a reset gate and an update gate. The reset gate decides how much past information to remember, and the update gate decides the ratio for the representation of the previous state and the current state. These processes are explained as Eqs. (1) to (4):

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (1)$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad (2)$$

$$\tilde{h}_t = \tanh(W x_t + r_t \odot U h_{t-1}) \quad (3)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \quad (4)$$

where *x* is the input data, *h* is the hidden state, and *W* and *U* are weights for the input data and the hidden state, respectively. At time step *t*, the results of the input variable that passed the reset gate and the update gate in the past hidden state are r_t and z_t, respectively. $\tilde{h}_t$ means information in memory to be used at the present time by determining the data to be forgotten from the past hidden state and adding this value to the current input data. h_t calculates the information in memory to finally be output by synthesizing the calculated values.

A base model was developed to compare prediction accuracies according to optimization technique. The base model consisted of an input layer, one GRU layer, and an output layer, and the number of neurons in the GRU layer was 50. The activation function was ReLU, the loss function was the mean squared error (MSE), and the learning rate was 1e-4. The batch size was set to 12, and of the number of epochs was set to 100. Additionally, the EarlyStopping callback was applied to prevent overfitting.

2.4 Hyperparameter optimization

Neural network models exhibit different predictive performances depending on the values of their hyperparameters. However, it is difficult to test all combinations of hyperparameters, and standardized tuning rules have not been determined. Grid search and random search methods are commonly used, but they have the disadvantage that the search takes a long time, and it is difficult to find the optimal combination of hyperparameters.

Therefore, in this study, hyperparameter tuning was performed by applying more advanced methods: Bayesian optimization and Hyperband optimization. Bayesian optimization probabilistically explores hyperparameter combinations based on a Gaussian process and performs an effective search by reflecting the previous results in the next search (Choi et al., 2022). However, it presents a problem in that tuning takes a long time, and overfitting may occur if the result is greatly affected by the hyperparameter values. Hyperband optimization is specialized for parallel computation, and it randomly extracts various combinations within a given hyperparameter range and finds the hyperparameter combination having the best performance based on the successive halving algorithm (SHA). Compared with Bayesian optimization, search time is saved, but because hyperparameter combinations are randomly extracted, the optimal value cannot be guaranteed when the range of combinations is large. In this study, the number of hidden layers, number of neurons, dropout rate, activation function, and learning rate were selected as hyperparameters for optimization. Table 1 presents the search range for each hyperparameter.

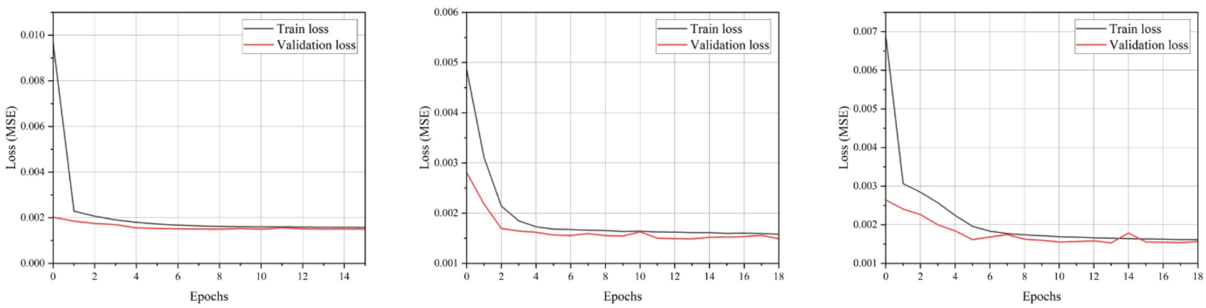
Table 1 Types of hyperparameters and ranges

Hyperparameters	**Ranges**
Number of hidden layers	0–3
Number of neurons	10–100, step = 2
Dropout rate	0–0.5, step: 0.1
Activation functions	tanh, sigmoid, ReLU
Learning rate	1e-10–1e-4

3. RESULTS AND DISCUSSIONS

3.1 Prediction accuracy of the occupancy prediction models

The base model (Case 1), Bayesian optimization (Case 2), and Hyperband optimization (Case 3) were used to train the model, and a performance evaluation was performed using the test data that were not used for training. The evaluation metrics were mean absolute error (MAE) and root mean squared error (RSME), which are based on the error between the actual value and the predicted value. The closer the two indicators are to 0, the better the prediction performance. Fig. 3 shows the learning state for each case. No overfitting occurred in any of the three models, and training was terminated without exceeding 20 epochs. This was due to the EarlyStopping callback; thus, training was done rapidly.



(A) Case 1 (B) Case 2 (C) Case 3

Fig. 3 Loss value for each case

Table 2 presents the combination of hyperparameters used for optimization, and the MAE and RMSE values measured on the test data. The optimal structure for Case 2 was 5 hidden layers, and the number of neurons per hidden layer was 10, 100, 10, 10, and 100. The dropout rate, activation function, and learning rate were set to 0, tanh, and 0.0001, respectively. The optimal structure for Case 3 consisted of three hidden layers with 86, 22, and 48 neurons, respectively. The dropout rate and activation function were identical to those used for Case 2, and the optimal learning rate was 0.0022.

Both the MAE and RMSE demonstrated outstanding prediction accuracy by all three models. Considering that the number of people included in the test data was approximately 30, an MAE of less than 1 indicates high prediction accuracy. Although the difference in the metric values for each model was insignificant, Case 1 achieved the lowest MAE of 0.8646 person, and Case 2 achieved the best RMSE of 1.7394 person. The RMSE was selected as the metric for determining the optimal predictive model. Because RMSE is calculated by taking the square root of MSE, if the absolute value of the error is large, the value is calculated proportionately. Therefore, a low RMSE indicates that relatively few large errors occurred, and that stable prediction is possible. Therefore, the optimal occupant prediction model was determined to be Case 2.

Table 2 Results of training

Cases	**Hyperparameters**				**Metrics**	
	Structure	**Dropout rate**	**Activation function**	**Learning rate**	**MAE**	**RMSE**
Case 1 (Base model)	6-50-1	0	ReLU	0.0001	0.8646 person	1.7560 person
Case 2 (Bayesian optimization model)	6-10-100-10-10-100-1	0	tanh	0.0001	0.8684 person	1.7394 person
Case 3 (Hyperband optimization model)	6-86-22-48-1	0	tanh	0.0022	0.9105 person	1.7753 person

Fig. 4 summarizes the comparison results for the actual value and the predicted value on the test data. Predicted values follow the actual values well overall, but according to the graph on the right, which shows time steps (5,500–6,000), a relatively large error occurred for Case 3 as compared with Case 1 and Case 2. In addition, by referring to the graph, it is possible to define problems that are commonly found in the three cases. First, the predicted value reflects the increase and decrease trend well, but oscillation occurs. Second, a shifting phenomenon occurs in which the t-1th actual value is used as the tth predicted value in the section where the number of occupants changes unexpectedly. As a result of various previous studies and data investigations, these two phenomena frequently occur when a recurrent neural network is used for time series prediction. This is because it is impossible to accurately predict the future if the future prediction target does not have special regularity and the probabilities of increase and decrease are the same. The occupant number data fits this description. Although the office shows a distinct pattern for commuting time and lunch break, the time at which an event occurs is not always fixed, and the occupant number at this time is also not constant. Because this corresponds to the natural characteristics of the prediction target, it is important to understand the limitations of the occupancy prediction model and apply it appropriately for the purpose of use.

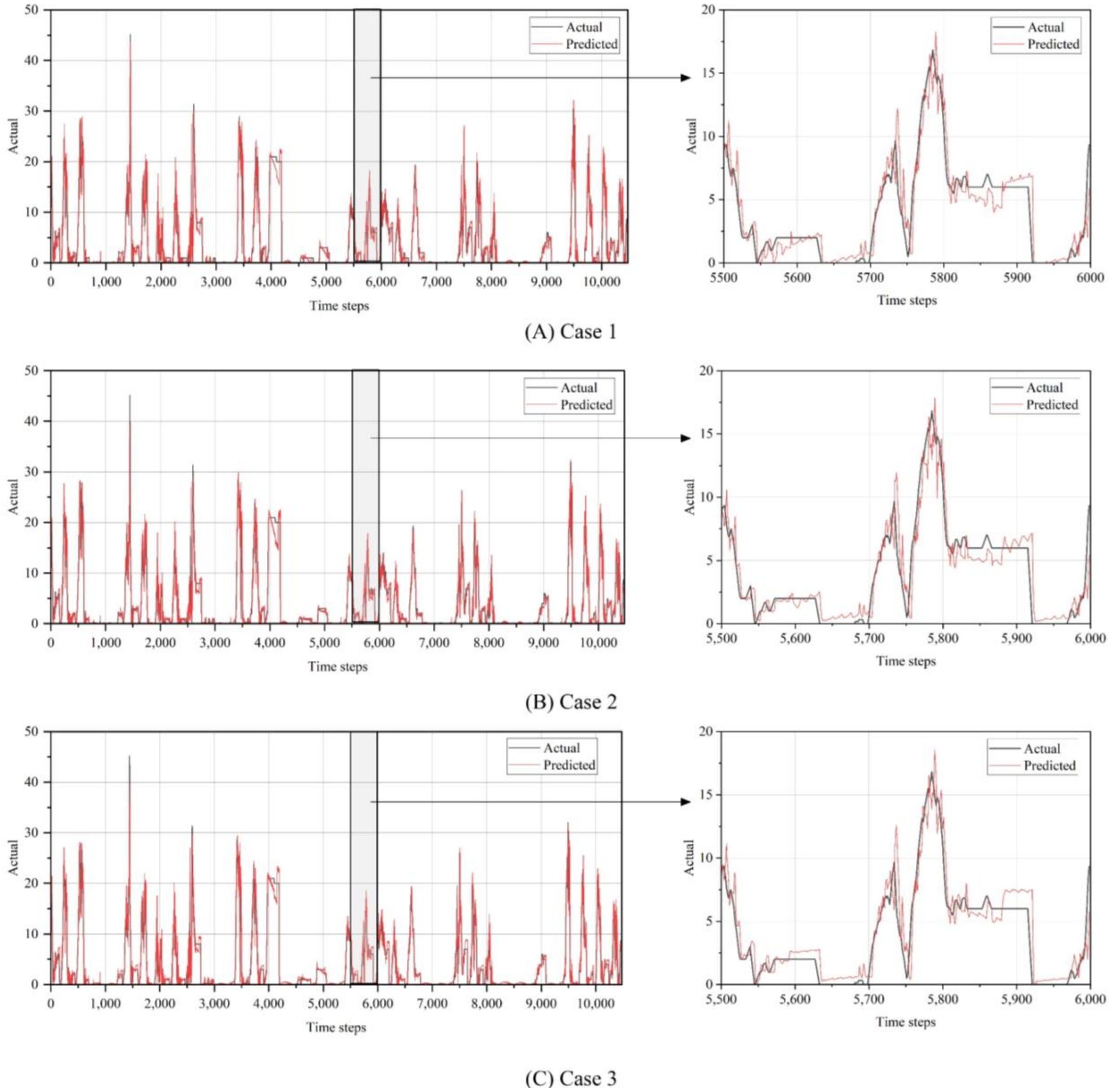


Fig. 4 Comparison of actual and predicted value

4. CONCLUSION

Recently, as interest in OCC has increased, the importance of occupant number prediction has been emphasized. Although various machine learning-based occupant number prediction models have been developed, most of them are focused on performance comparison between machine learning models. However, there has not been enough research done on performance improvement by optimization. Therefore, the purpose of this study is to apply various optimization techniques in the development stage of the occupant number prediction model and to derive the optimal model. Additionally, the limitations of the occupant number prediction model and future research directions are mentioned.

GRU was used as the machine learning model, and three models were developed. The base model, Bayesian optimization, and Hyperband optimization models were respectively applied to optimize the hyperparameters. The performance evaluation on the test data showed that the base model had the best MAE of 0.8646 person, and the model to which Bayesian optimization was applied showed the best RMSE of 1.7394 person. Finally, the model to which Bayesian optimization was applied was selected as the optimal model, considering its prediction stability.

All three models presented outstanding prediction accuracy. However, although the hyperparameter optimization was applied, the performance differences of the three models were insignificant. It was found that, the reason was due to the uncertainty characteristic of the occupant number. Also, a shifting phenomenon occurred intermittently, which was also considered to be due to the uncertainty characteristics. Therefore, in future work, alternative proposals that would address these limitations, and research on the development of control algorithms should be conducted.

Acknowledgements

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of the Republic of Korea (No. 20212020800120).

5. REFERENCES

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches, *arXiv preprint arXiv*, 1409-1259.

Choi, E. J., Park, B. R., Kim, N. H., & Moon, J. W. (2022). Effects of thermal comfort-driven control based on real-time clothing insulation estimated using an image-processing model, *Building and Environment*, 223, 109438.

Choi, Y. J., Park, B. R., Hyun, J. Y., & Moon, J. W. (2022). Development of an adaptive artificial neural network model and optimal control algorithm for a data center cyber–physical system, *Building and Environment*, 210, 108704.

Huchuk, B., Sanner, S., & O'Brien, W. (2019). Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data, *Building and Environment*, 160, 106177.

Kim, S., Kang, S., Ryu, K. R., & Song, G. (2019). Real-time occupancy prediction in a large exhibition hall using deep learning approach, *Energy and Buildings*, 199, 216-222.

Luo, N., Wang, Z., Blum, D., Weyandt, C., Bourassa, N., Piette, M. A., & Hong, T. (2022). A three-year dataset supporting research on building energy management and occupancy analytics, *Scientific Data*, 9(1), 1-15.

Panchabikesan, K., Haghighat, F., & El Mankibi, M. (2021). Data driven occupancy information for energy simulation and energy use assessment in residential buildings, *Energy*, 218, 119539.

Park, J. Y., Ouf, M. M., Gunay, B., Peng, Y., O'Brien, W., Kjærgaard, M. B., & Nagy, Z. (2019). A critical review of field implementations of occupant-centric building controls, *Building and Environment*, 165, 106351.

Schiele, J., Koperna, T., & Brunner, J. O. (2021). Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks, *Naval Research Logistics (NRL)*, 68(1), 65-88.