Characterising Window Opening Behaviour of Occupants Using Machine Learning Models

Bongchan Jeong[1], Heewon Choi[2], Jihyun Yoo[2], and J.S. Park[*3]

Industry-University Cooperation Foundation, Hanyang Univversity
222 Wangsimni-ro
Seongdong-gu, 133-791, Seoul, 133-791, South Korea

Department of Architectural Engineering, Graduate school of Hanyang Univversity
222 Wangsimni-ro
Seongdong-gu, 133-791, Seoul, 133-791, South Korea

Department of Architectural Engineering, College of Engineering, Hanyang University
222 Wangsimni-ro
Seongdong-gu, 133-791, Seoul, 133-791, South Korea
**Corresponding author: junpark@hanyang.ac.kr*

ABSTRACT

Occupants control indoor environments to meet their individual needs for comfort. The control of window is the most common natural ventilation method influencing indoor environment as well as the energy use of the buildings to maintain a suitable environment. Therefore a better understanding of window control behaviour of the occupants has significant implication to enhance occupant comfort with minimal energy consumption. The objective of this study was to identify an appropriate algorithm and variables to develop a predictive model for window control. A longitudinal field measurement was performed for 10 months in 23 residential houses. Outdoor and indoor environmental conditions and window status were continuously monitored for the period. To identify an appropriate modelling algorithm, the logistic regression which is a traditional statistical method for binary data and three popular machine learning models, k-Nearest Neighbours (KNN), Random Forest (RF) and Artificial Neural Networks (ANN) were applied and compared. The result of this study reveals that the machine learning algorithms outperforms the traditional statistical regression model. Another contribution of this study is that variables influencing occupants to control window were varied in each season and from person to person. Thus, these results show the improvement of prediction with the use of machine learning-based control system.

KEYWORDS

Occupant behaviour, Indoor air quality, Window manual control, Machine learning, Energy simulation

1 INTRODUCTION

Occupants control surrounding environments to meet their individual needs for comfort. The control of windows is the most common method for ventilation and thermal comfort purpose. The fresh air supplied through window openings can improve indoor air quality by diluting indoor air contaminant (Park, 2013). It can also improve thermal comfort by dropping indoor air temperature and by encouraging air flow (Raja, 2001). Therefore energy consumption can vary due to such different occupant behaviour (Park et al., 2012; Jeong et al., 2016). Therefore a better understanding of window control behaviour of the occupants has significant implication to enhance occupant comfort with minimal energy consumption.

This study aims to identify an appropriate algorithm and primary predictors for predicting window opening behaviour in residential buildings. Based on a longitudinal field measurement of twenty three residential housing units in Korea, predictive accuracy of a logistic regression model which is a traditional statistical method for binary data and three popular machine learning models, k-Nearest Neighbours (KNN), Random Forest (RF) and Artificial Neural Networks (ANN) were compared.

2 METHOD

2.1 Samples

Table 1 describes the twenty three housing units who participated the field measurement. The subjects were selected from three multi-family residential complexes that are located in suburban areas of Seoul in Korea. All units had a hydronic radiant heating floor system and at least one air-conditioner for cooling.

Table 1: Characteristics of the sample housing units

Sample ID	Floor area [m^2]	Floor level[a]	Number of occupants[b]	Period of residence [year]	Smokers
a	109	4/20	3 (M, F, m)	6.5	0
b	109	8/20	4 (M, F, f, f)	6.8	0
c	171	8/20	4 (M, F, f, f)	6.6	1
d	109	16/24	4 (M, F, f, f)	7.0	0
e	109	5/24	3 (M, F, f)	6.6	0
f	129	14/26	4 (M, F, m, f)	7.0	0
g	163	12/14	4 (M, F, m, f)	2.3	0
h	163	4/15	4 (M, F, m, f)	2.0	0
i	163	6/14	4 (M, F, m, f)	2.3	1
j	163	10/15	4 (M, M, F, F)	2.0	0
k	163	5/14	4 (M, F, m, f)	2.4	1
l	136	2/19	4 (M, F, m, m)	2.8	0
m	145	4/19	4 (M, F, f, m)	3.0	0
n	145	13/19	4 (M, F, f, f)	1.2	0
o	163	1/19	5 (M, F, m, f, f)	2.8	0
p	72	14/21	4 (M, F, f, f)	6.7	0
q	80	7/23	4 (M, F, m, f)	9.5	1
r	72	14/21	3 (M, F, M)	7.8	0
s	108	9/25	4 (M, F, f, f)	4.5	0
t	79	25/25	4 (M, F, m, f)	17.5	0
u	79	6/20	4 (M, F, m, m)	1.4	0
v	163	15/25	4 (M, F, m, f)	2.0	0
w	79	19/25	4 (M, F, m, m)	16	0

[a]Floor number/Total Number of floors.
[b]M, male adult; F, female adult; m, male below 18 years; f, female below 18 years.
[c]H, heating period; NH, non-heating period; C, cooling period

Table 2: Measurement parameters and the specifications of devices used for measurement

Parameter		Device	Accuracy	Interval [minutes]
Outdoor	Dry-bulb temperature Relative humidity	Data Logger[a] (TR-72ui and RS-11, T&D)	±0.3 ℃ ±5% RH	10
Indoor	Dry-bulb temperature Relative humidity CO_2 concentration	Humidity, Temperature and CO_2 monitor (MCH-383SD, Lutron)	±0.8 ℃ ±4% RH ±40 ppm	10
	Mean radiant temperature	Data Logger (UX100-014M, Onset Computer)		
Window opening	Window opening status	State Logger[b] (UX90-001, Onset Computer)	-	Event

[a]The sensors were radiation-shielded.
[b]Binary window status was measured, opened or closed

2.2 Measurements

The field measurement were conducted on three separate periods: Non-heating period: from February to May 2014; Heating period: From December 2014 to February 2015; Cooling period: From May to September 2015. Outdoor and indoor environmental parameters were continuously measure at 10 minute intervals.

2.3 Logistic regression model

We used a logistic regression model and three machine learning algorithms. A logistic regression analysis is a traditional statistical method for predicting binary data. A logit probability distribution can be expressed as follows:

$$\mathrm{p}(x_1, \dots, x_n) = \frac{\exp(\alpha+\beta_1 x_1+\cdots+\beta_n x_n)}{1+\exp(\alpha+\beta_1 x_1+\cdots+\beta_n x_n)} \quad (1)$$

where p is a probability of windows to be opened, $x_1,\dots,x_n$ are predictors, α is a constant, $\beta_1,\dots,\beta_n$ are coefficients estimated using maximum likelihood estimation. Therefore the equation calculates window opening probability with given conditions of predictors.

2.4 Machine learning models

The three machine learning algorithms used in this paper are K-Nearest Neighbors (KNN), Random Forest (RF) and artificial neural network (ANN). KNN stores the data set as training data and classifies a new data based on a similarity measure. We set the Number of K as automatic between 3 and 5 and Euclidean distance function was used. RF produces mean prediction from many decision trees of the data set. We used 10000 nodes with 10 maximum tree depth. ANN has three different layers and adjusts connection weights and bias of neurons within the layers. We used feedforward neural network with 4 hidden neurons.

3 RESULT

3.1 Predictive performance

Predictive accuracy of the algorithms were compared to identify an appropriate algorithm for predicting window opening behaviour. CV(RMSE) was used for accuracy comparison. The lower CV(RMSE) value, the better predictive performance it is. Table 3 summarizes the CV(RMSE) of the four algorithms for the 23 subjects. It shows that KNN displayed the best predictive performance among the algorithms. KNN's average CV(RMSE) is the lowest among

Table 3: Predictive accuracy comparison of KNN, RF, ANN and LR using CV(RMSE). Values in bold indicate the best performance algorithm for each subject.

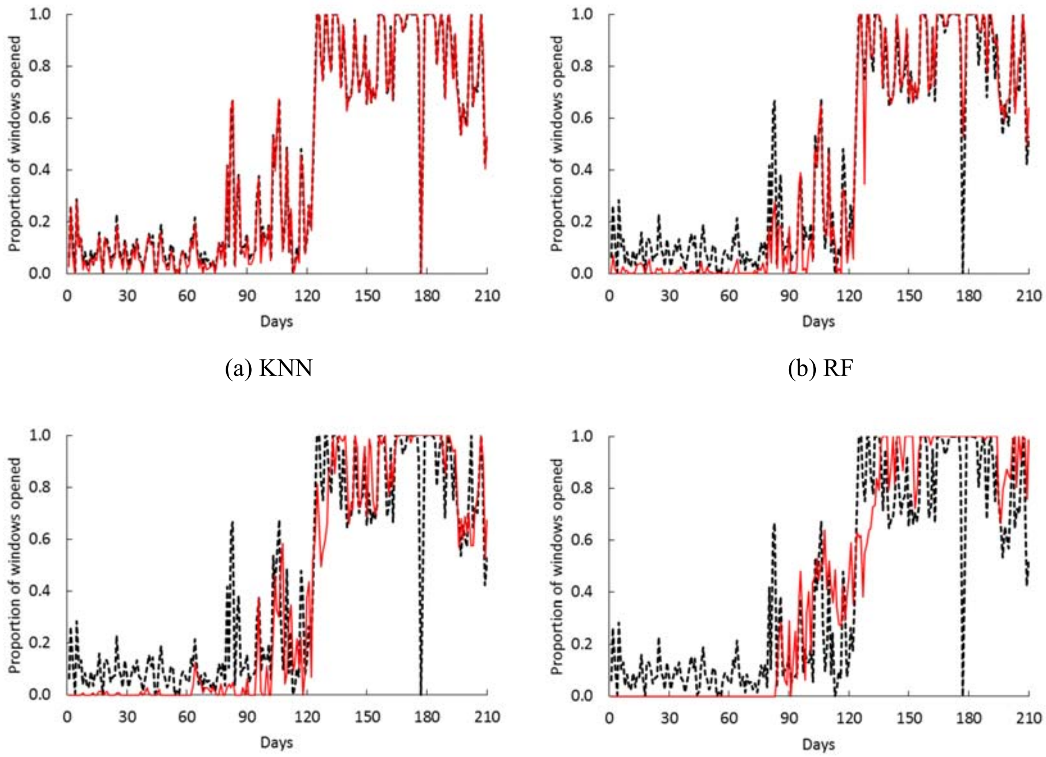
Subjects	kNN	RF	ANN	LR
a	0.022	**0.012**	0.062	0.045
b	**0.034**	0.050	0.122	0.123
c	**0.034**	0.063	0.161	0.211
d	**0.009**	0.026	0.080	0.096
e	0.026	**0.026**	0.073	0.094
f	**0.024**	0.036	0.083	0.081
g	**0.004**	0.015	0.043	0.042
h	**0.007**	0.021	0.086	0.080
i	**0.003**	0.030	0.145	0.160
j	**0.009**	0.028	0.101	0.118
k	**0.005**	0.011	0.077	0.074
l	**0.005**	0.005	0.010	0.010
m	**0.007**	0.010	0.069	0.077
n	**0.015**	0.023	0.119	0.139
o	**0.006**	0.038	0.189	0.227
p	**0.011**	0.029	0.096	0.097
q	**0.006**	0.011	0.158	0.184
r	**0.013**	0.022	0.061	0.059
s	**0.007**	0.025	0.071	0.069
t	**0.018**	0.068	0.142	0.158
u	**0.015**	0.045	0.130	0.133
v	**0.001**	0.008	0.102	0.330
w	**0.014**	0.018	0.020	0.020
Mean	0.013	0.027	0.096	0.114

the algorithms, followed by RF, ANN and LR. KNN displayed the best performance from 21 subjects and RF displayed the rest 2 subjects. It indicates that machine learning algorithms outperforms a logistic regression algorithm and KNN showed the best performance among the tested algorithms in this paper.

Figure 1 shows predictive performance of the algorithms using the daily proportion of windows opened. The proportion of windows opened is the number of data that windows are opened divided by the total number of data. KNN's prediction showed the best fit among the algorithms. RF and LR showed less accurate result especially during the heating period. LR was showed the poorest performance and it failed to predict window openings for the first 80 days in the heating period.

3.2 Variable importance

Optimum variables for model development has to be identified to lower data collection cost and computational cost. We used Gini importance which can be used as an indicator for feature relevance. Standardized Gini importance was used to compare relative relevance of parameters. Figure 2 shows the standardized Gini importance result for each subject. We grouped parameters into four groups: Temp_out includes outdoor temperature-related parameters such as daily average outdoor temperature; Temp_in includes indoor temperature-related parameters; Temp_diff includes temperature difference between outdoor and indoor environment; Env_out includes outdoor environmental parameters; Env_in includes indoor environmental factors; Time includes time series parameter. The result shows that temperature-related parameters are the primary parameters that are required to be used as predictors.



(a) KNN (b) RF

(a) ANN (b) LR

Figure 1: Comparison of KNN's RF's ANN's and LR's predicted proportion of windows opened with observed data. The dashed line represents observed data and the solid red line represents predicted data.

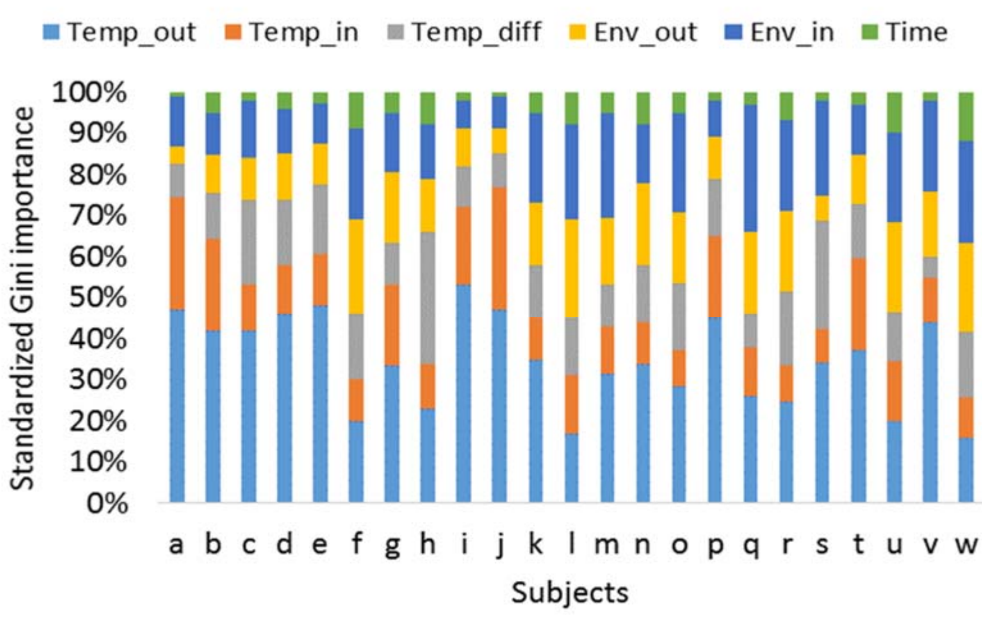


Figure 2: Variable importance for each subject.
(Standardized Gini importance was used for the indicator)

4 DISCUSSION

We compared predictive performance of a logistic regression model which is a traditional statistical method and the three machine learning algorithms. As a result, KNN showed the best performance among the tested algorithms. We only focused on predictive accuracy while there are other features that needs to be considered for model development. As KNN stores and uses the whole data set at each prediction. The larger the data set is, the more it becomes computationally expensive. Therefore costs for computation and data collecting have to be also considered for the model development.

We used standardized Gini importance for evaluating variable importance. However there are many other method. For example, stepwise method which is comparing performance of an algorithm by adding variables until it includes the all variables. Other factors that affect the quality of models can also be considered such as robustness. Further development regarding model development will be reported.

5 CONCLUSION

This study aimed to identify an appropriate algorithm and primary parameters for predicting window opening behaviour in residential buildings. We compared predictive accuracy of a logistic regression model and the three machine learning algorithms. The tested machine learning algorithms outperformed the logistic regression model. KNN showed the best performance but it requires to consider its computational cost. Temperature-related variables are the major predictors to be used for window opening prediction. These results show improvement of predictive accuracy with the use of machine learning-based control system.

6 ACKNOWLEDGEMENTS

This work was supported by National Research Foundation of Korea (NRF) with a grant funded by the Korean government (MEST) (NRF- 2017R1A2B3009344).

7 REFERENCES

Raja, I.A., Nicol, J.F., McCartney K.J. and Humphreys M.A. (2001). Thermal comfort : use of controls in naturally ventilated buildings. *Energy and Buildings,* 33, 235-244.

Park, J.S. (2013). Long-term field measurement on effects of wind speed and directional fluctuation on wind-driven cross ventilation in a mock-up building. *Building and Environment,* 62, 1-8.

Park, J. S. and Kim, H. J. (2012). A field study of occupant behavior and energy consumption in apartments with mechanical ventilation. *Energy and Buildings*, 50, 19-25.

Jeong, B., Jeong J. and Park J.S. (2016). Occupant behavior regarding the manual control of windows in residential buildings, *Energy and Buildings*, 127, 206-216.