Abstract

Fanger's PMV index has been studied by many researchers as a control parameter for maintaining comfortable indoor environment using HVAC systems. Recently, couple of data-driven modelling techniques have been used for the prediction of thermal comfort. Machine learning-based data-driven models for thermal comfort can predict the PMV with indoor parameters and human factors. However, current models still have two limitations. The first is that the input variables used for thermal comfort prediction still utilize Fanger’s six variables, which are difficult to measure in real buildings. Hence, it is necessary to develop a data-driven model for predicting the PMV using data that can be easily monitored in real buildings. The second limitation is uncertainty associated to the predicted value of thermal comfort values from a data-driven model. In particular, uncertainty must be considered in real buildings, because the monitoring data from various sensors or measuring instruments are affected by noise or other sources, and lead to errors in the results. Therefore, this study aims to develop a Gaussian process regression (GPR) model for the prediction of thermal comfort, including uncertainty information. As a results, we identified the influences of the environmental factors on the thermal comfort of an occupant in a room and employed them to develop a data driven model using the GPR. The results of the thermal comfort models clearly show that the relative humidity and ambient temperature are major variables in enhancing the accuracy of the prediction. In addition, the 95% confidence interval was also significantly narrowed, indicating that the model uncertainty was reduced.

Introduction

In recent years, as the importance of energy consumption has increased, many building-related data have been measured and collected to maintain comfort and low energy consumption. Using these data, researchers have been able to predict thermal comfort under diverse conditions that determine thermal sensation. Especially, Fanger’s PMV index has been studied by many researchers as a controlled parameter to accommodate the control of HVAC systems. Recently, couple of data-driven modelling techniques have been used for the prediction of thermal comfort. For example, Zolfaghari et al., Guan et al., and Hoppe developed a new model to use the non-steady and transient environments to characterize thermal comfort using data-driven methods (Zolfaghari et al., 2010; Guan et al., 2003; Hoppe, 2002). In another study, owing to their ability to approximate non-linear model functions, neural networks are widely used PMV and the accessible variable pair mapping of qualitative relationships (Grabe, 2016; Songuppakarn et al., 2014). They found that the predicted PMV values agreed well with those obtained using the model proposed by Fanger and questionnaires. However, these approaches still have two limitations. The first is that the input variables used for thermal comfort prediction still utilize Fanger’s six variables (i.e., air temperature, air humidity, air flow rate, and average radiation temperature, clothing thermal resistance, metabolic rate), which are difficult to measure in real buildings. For example, the MRT is commonly determined using a black globe temperature device, which consists of a painted black sphere with a temperature-sensing device at its center. This temperature-sensing device is not commonly employed in buildings. Hence, it is necessary to develop data-driven models for predicting the PMV using data that can be easily monitored in a building. The second limitation is the uncertainty of the predicted value of thermal comfort values from a data-driven model. In particular, uncertainty must be considered in real buildings, because monitoring data obtained through various sensors or measuring instruments are affected by noise, and various errors may occur.

Recently, owing to its characteristic of quantifying uncertainty in predictions, Gaussian process regression (GPR) has been widely used to predict existing systems and energy usage in the building sector. In other words, GPR can provide predictive values in the form of probabilistic distributions. Therefore, the Gaussian modelling results reflect the prediction uncertainties, which cannot be quantified directly using neural networks (Neal, 1995). Hence, GPR can be useful as a statistical modelling method in research on predictions that consider parameter uncertainty and parametric variability. So, a GPR approach is proposed in this study for thermal comfort indicator prediction, to enable the uncertainty in the thermal comfort estimate to be quantified in the form of a probability distribution.

The main purpose of this study is to identify the influence of the environmental factors on the thermal comfort in buildings, thereby developing a data-driven model using the Gaussian process.
Methods

For this study, first, an experimental study in a test chamber was conducted to develop a thermal comfort model using GPR. In this step, we identified the influence of the environmental factors (e.g., globe temperature, ambient temperature, MRT, relative humidity, and air speed) on the occupant’s thermal comfort in a room, thereby developing a data-driven model using the GPR. Then, based on the experimental results in the chamber, we applied this model to a real building and analysed the results to assess its applicability.

Data acquisition

The data collection for PMV prediction model was conducted in a “living test bed” that provided an environment similar to those found in residential buildings, as shown in Figure 1. The testbed is an enclosed chamber that consists of variable refrigerant flow (VRF) system, a heat recovery ventilator, a humidifier, a dehumidifier, a floor heating system, a lighting system with dimming control, an automatic blind, and a building monitoring and control system. We collected the five measured variables (e.g., globe temperature, ambient temperature, MRT, relative humidity, and air speed). For this, as shown in Table 1, a data logger was positioned in the middle of the test bed to collect the data. This measuring device contains various sensors for collecting environmental data, such as the globe temperature ($T_g$), air temperature ($T_a$), RH, and air speed ($V_a$). The measurement accuracy of $T_g$ and $T_a$ is less than ±0.2 °C from -10 °C to 100 °C, that of RH is ±2.5% from 5% to 98%, and that of $V_a$ is ±0.05 m/s from 0 m/s to 1 m/s. To measure and assess the six parameters related to PMV, the accuracies required by ISO 7726 and in this study are summarized in Table 1. Then, the MRT, i.e., $T_{r}$, can be calculated using $T_g$, $T_a$, and $V_a$ by applying the following formula ISO 7726.

The occupant-related values of the thermal resistance values, i.e., $I_{cl}$ and MET, can be entered directly according to the subjective state based on the “thermal resistance values” and “metabolic rate” in ISO 9920 (2007). In this study, $I_{cl}$ was set to 0.7 corresponding to daily clothing (i.e., underwear, shirts, trousers, socks, and shoes) and MET was set to 1.2 (comfortably seated: light manual work such as writing, typing, drawing, cutting, or accounting). In reality, for most cooling and heating control applications, the environment is controlled in a closed space, and most of the occupants can be assumed to be dressed similarly. In addition, it is likely that they will be performing at similar activity levels. Then, PMV was calculated according to ISO 7730 (1993). Finally, the GPR-based PMV prediction model developed in this study was compared with the calculated PMV value.

The test bed air temperature was controlled by a VRF system to maintain the set point. The initial set-point temperature was 19 °C, which was held for 90 minutes to ensure that the temperature in the test bed was stabilized. The temperature was then increased to 30 °C for 50 minutes and subsequently returned to 19 °C. The air volume (or flow rate) of the VRF was set to automatic control mode.

Table 1: The collected data used for PMV prediction

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Accuracy</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_g$(Globe temperature)</td>
<td>±0.1 °C</td>
<td>Measurement</td>
</tr>
<tr>
<td>$T_a$(Ambient temperature)</td>
<td>±0.1 °C</td>
<td>Measurement</td>
</tr>
<tr>
<td>$T_r$ (MRT)</td>
<td>-</td>
<td>calculation</td>
</tr>
<tr>
<td>$V_a$ (Air speed)</td>
<td>±0.05 m/s</td>
<td>Measurement</td>
</tr>
<tr>
<td>RH</td>
<td>±2.5 %</td>
<td>Measurement</td>
</tr>
<tr>
<td>$I_{cl}$</td>
<td>-</td>
<td>specific value (0.7)</td>
</tr>
<tr>
<td>MET</td>
<td>-</td>
<td>specific value (1.2)</td>
</tr>
</tbody>
</table>

Composition data set and evaluation

Before organizing the data sets for the thermal comfort predictive model, as shown in Table 1, the data set was constructed by selecting variables (i.e., $T_g$, $RH$, $V_a$) that are generally easily measurable among the various parameters required for PMV calculation. And six predictor sets were constructed based on the selected main informative predictors and evaluated. The data collected through the abovementioned process were randomly extracted, and the model was learned by constructing a training set with 80% of the data and a test set with 20% of the data. The predictive power of the GPR model was evaluated using four statistical quantities: (1) mean bias error (MBE); (2) root-mean-square error (RMSE); (3) total sum-of-squares error (SSE); (4) $R^2$ measures the proportion of total variation explained by the fitted regression model.

Gaussian Process Regression for thermal comfort

This section explains the Gaussian process regression used as the thermal comfort prediction model in this study. The basic principles of the Gaussian process and the mathematics of this section are based on the work of Rasmussen and Williams (2006). A key advantage of the Gaussian process is that it constructs the model by specifying the structure of the covariance matrix of the explanatory variables, rather than the algebraic structure of the input–output relationship itself, as it is done in traditional parametric regression approaches (Rasmussen and Williams, 2006; Zhang et al., 2013). We have a training set $\{(x_i, y_i); i = 1, 2, ..., n\}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, drawn from an unknown distribution. A Gaussian process model addresses the question of predicting the value of a response variable $y_{new}$ given the new input vector $x_{new}$ and the training data. Gaussian process is completely specified by its mean function $m(x)$ and...
covariance function, \( k(x, x') \) of a real process \( f(x) \) as (That is, if \( f(x), x \in \mathbb{R}^d \) is a Gaussian process, then \( E(f(x)) = m(x) \) and \( \text{Cov} \{ f(x), f(x') \} = E \{ (f(x) - m(x))(f(x') - m(x')) \} = k(x, x') \).)

\[
m(x) = E(f(x)) \tag{1}
\]

\[
k(x, x') = E \{ (f(x) - m(x))(f(x') - m(x')) \} \tag{2}
\]

And will write the Gaussian process as
\[
f(x) \sim \text{GP} (0, k(x, x')) \tag{3}
\]

That is, \( f(x) \) is from a zero mean Gaussian process with covariance function, \( k(x, x') \), which is also known as the kernel function. A variety of kernel functions, such as linear, exponential, rational quadratic, Matérn, and squared exponential have been used in different applications. Among these, the squared exponential is the most widely-used kernel in the field of kernel machines (Rasmussen and Williams, 2006). In this study, the choice of the covariance function is a squared exponential kernel, which takes the form:
\[
k(x, x') = \sigma_f^2 \exp\left[-\frac{1}{2} \frac{(x-x')^2}{\sigma_x^2}\right] \tag{4}
\]

And we used the ‘package GPFDA’ in R, which is a language and environment for statistical computing and graphics.

**Results**

Figure 2 shows the temperature measured by the sensors installed in the chamber, set-point temperature \( (T_{37}) \), and dry-bulb temperature \( (T_a) \) during the experiment.

![Figure 2: Measured temperature](image)

**Figure 2: Measured temperature**

In Figure 2, the red line indicates the time history of the set-point temperature of the test bed used in the experiment. The \( T_a \) patterns were changed to “low-high-low” according to the change of the set temperature. In reality, room temperature sensors are usually inside thermostats or cooling systems, and most of these devices are located on the walls or in the corners of rooms. In this study, although the value varied depending on the position of the temperature sensor, the temperature variations according to the sensor position were not considered. Figure 3 shows the RH pattern and thermal comfort indicator (i.e., PMV) during the experiment. At this time, the PMV value shows the value calculated by ISO 7730 based on the measured data as described in the previous section. In Figure 4, the variation of \( V_a \) with time is irregular with time, and it can be seen that \( V_a \) increases, especially when the set temperature changes. As shown in Fig. 2 ~ 4, we can confirm that the three variables \( (V_a, T_a, \text{ and } RH) \) are easily monitored variables in the building, but they show a certain pattern with correlation as the PMV changes.

**Development of thermal comfort prediction model**

Among all of the variables related to thermal comfort, we selected three input variables: \( V_a, T_a, \text{ and } RH \). Then, Gaussian process model for thermal comfort prediction were constructed using all possible combinations of these three variables, and the statistical characteristics of each set were evaluated, as presented in Table 2. The output was a thermal comfort indicator (i.e., PMV) calculated according to ISO 7730. Seven predictor sets were evaluated to determine the most effective method for obtaining an accurate thermal comfort indicator.

Table 2 lists the values of RMSE, MBE, and SSE. First, SSE is approximately 1.005 for Set 1, which was based on \( V_a \). With the three variables combined (Set 7), SSE reaches its lowest value, which is approximately 0.003, and with two variables (i.e., RH and \( T_a \)), SSE is also quite low. And MBE values are also very small, and close to zero.

**Table 2: Results of thermal comfort prediction using GPR with sets of predictors**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Variables</th>
<th>RMSE</th>
<th>MBE</th>
<th>SSE mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>( V_a )</td>
<td>1.025</td>
<td>-1.013</td>
<td>1.005</td>
</tr>
<tr>
<td>Set 2</td>
<td>RH</td>
<td>0.452</td>
<td>-0.726</td>
<td>0.026</td>
</tr>
<tr>
<td>Set 3</td>
<td>( T_a )</td>
<td>0.321</td>
<td>-0.212</td>
<td>0.068</td>
</tr>
<tr>
<td>Set 4</td>
<td>( V_a, RH )</td>
<td>0.451</td>
<td>-0.778</td>
<td>0.023</td>
</tr>
<tr>
<td>Set 5</td>
<td>( V_a, T_a )</td>
<td>0.274</td>
<td>-0.333</td>
<td>0.049</td>
</tr>
<tr>
<td>Set 6</td>
<td>( T_a, RH )</td>
<td>0.084</td>
<td>-0.057</td>
<td>0.007</td>
</tr>
<tr>
<td>Set 7</td>
<td>( V_a, T_a, RH )</td>
<td>0.062</td>
<td>-0.062</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Figure 5 show analysis results on the predictive models measured through CVRMSE and \( R^2 \). RMSE is a measure of accuracy that compares the differences between the values predicted by a model and those actually observed (Hyndman and Koehler, 2006). \( R^2 \) indicates how well the regression line approximates the actual data points. In general, \( R^2 \) is between 0 and 1, with \( R^2 > 0.8 \) indicating ideal values. In Figure 5, \( RMSE \) and \( R^2 \) for the set based on \( V_a \) (Set 1) are approximately 1.025 and 0.302, respectively. For Set 2, which included \( RH \), \( RMSE \) is decreased by approximately 0.452, and \( R^2 \) is improved to...
more than 0.86. RMSE decreases sequentially from 0.452 to 0.320, 0.273, and 0.804, finally reaching 0.062. With the three variables combined, RMSE reaches its lowest value. In particular, the high accuracy occurs in Sets 6 and 7, with RH and $T_a$ included. In these cases, RMSE is less than approximately 0.08, and $R^2$ exceeds 0.995.

Figure 5: $R^2$ and RMSE of GPR models with sets of predictors

The results of the thermal comfort models clearly show that the relative humidity and ambient temperature are variables that can substantially increase the accuracy. That is, the model accuracy depends on the inclusion of two variables: the relative humidity and ambient temperature. In addition, if the air velocity is used as the only variable, it is almost impossible to predict the thermal comfort. Nonetheless, the addition of the air velocity to the model (Set 7) including the temperature and humidity improved the accuracy compared to that of Set 6. Figure 6 present comparisons of the actual (observed) values for thermal comfort (i.e., PMV) with those predicted using the GPR models.

The x axis corresponds to time (time step: 1 minute), and the y axis indicates thermal comfort. The red circles indicate the actual thermal comfort, while the blue lines show the predicted thermal comfort obtained by performing GPR (i.e., the mean of the posterior distribution). The blue boundary shows the 95% confidence interval, i.e., two standard deviations on either side of the mean, since the confidence interval for the predicted value can provide uncertainty information for the model prediction. The fact that GPR provides the distribution of the output instead of a point value is a key advantage. Figure 6, which corresponds to the model with the air velocity as the only variable (i.e., Set 1), shows that the predicted and actual values are not close to one another, and the 95% confidence interval is very broad. In other words, the uncertainty of the predicted value is very large. However, the results for Sets 2–5, each of which included either relative humidity or air temperature, exhibit confidence intervals narrower than that of Set 1. At the point at which PMV suddenly decreases (e.g., the 50th point in Set 2), the prediction performance is somewhat lower, but in the remaining intervals, the predicted values and actual values are almost the same. However, for Set 6, which included both relative humidity and air temperature, the prediction is also good at this point.

Figure 6: Prediction results under uncertainty
In addition, the 95% confidence interval is significantly narrowed, indicating reduced model uncertainty. For Set 7, with three parameters as variables, including the air velocity, the 95% confidence interval is narrower and the uncertainty is very low. In other words, the accuracy and reliability of the model are affected by the quality of the training data and the training variables.

Figure 7: Error histogram

In Figure 7, we compared an error density between the observed and predicted values with the Gaussian process. These results are similar to those described above. The error histogram for Set 1 has a wide range of values from -3 to 3. Meanwhile, Sets 6 and 7 show lower error ranges (i.e., their values have ranges of ±0.2).

Application of thermal comfort prediction model in an actual building

The experimental results presented in the previous section show that the temperature and humidity (and air velocity, if possible) can be used to predict the thermal comfort of occupants. In addition, the GPR model proposed in this report enables predictions to be made with uncertainty information in the form of a confidence interval. In this study, it was assumed that all changing environmental conditions are reflected in various environmental factors such as air temperature (when cooling, heating, or using solar radiation), relative humidity, and air velocity (when opening windows or ventilating). However, the experiments described above were conducted in a closed chamber without influencing factors such as solar radiation or ambient temperature. Therefore an experimental study was performed in an actual building affected by solar radiation or ambient temperature to verify the test results presented in the previous section. The test building was the engineering building located in a Dankook university campus in Yongin, Korea. The building consists of offices and seminar rooms, as shown in Figure 8.

Table 3: statistical analysis results

<table>
<thead>
<tr>
<th>Data set</th>
<th>Variable</th>
<th>Variance</th>
<th>Min</th>
<th>Max</th>
<th>SD.</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Data</td>
<td>$T_a$</td>
<td>8.13</td>
<td>17.06</td>
<td>29.30</td>
<td>2.85</td>
<td>22.45</td>
</tr>
<tr>
<td></td>
<td>RH</td>
<td>26.96</td>
<td>12.3</td>
<td>42</td>
<td>5.19</td>
<td>22.84</td>
</tr>
<tr>
<td></td>
<td>PMV</td>
<td>0.49</td>
<td>-2.13</td>
<td>1.11</td>
<td>0.07</td>
<td>-0.46</td>
</tr>
<tr>
<td>Test Data</td>
<td>$T_a$</td>
<td>0.38</td>
<td>21.8</td>
<td>25</td>
<td>0.62</td>
<td>22.93</td>
</tr>
<tr>
<td></td>
<td>RH</td>
<td>9.57</td>
<td>34.2</td>
<td>44.9</td>
<td>3.09</td>
<td>40.17</td>
</tr>
<tr>
<td></td>
<td>PMV</td>
<td>0.02</td>
<td>-0.53</td>
<td>0.35</td>
<td>0.16</td>
<td>-0.21</td>
</tr>
</tbody>
</table>

Figure 9: Result of thermal comfort prediction in room B

The training data were collected for eight days from March 1 until March 12, 2018, in 3F room A of this building. In addition, test data were collected from May 4 until May 10 in the room B. The temperature, the relative humidity, and PMV were recorded at 10 min intervals during the experiment periods. The indoor conditions of the two test rooms had different temperature and humidity ranges, owing to the different measurement periods and different occupants. To verify that the thermal comfort of occupants can be predicted by using the temperature and humidity under changing real environmental conditions, we used two data items measured at different locations and during different periods. Based on the training data, we constructed a GP prediction model using the temperature and humidity and analyzed the prediction accuracy by applying the test data. Table 3 summarizes the statistical quantities for these collected data. Prediction using GPR requires the training set to cover most of the input domain (Yan, 2013). It can be confirmed that the temperature, humidity, and PMV ranges of the training data used to develop the predictive model include the ranges of the test data.
experiment described in the above section. In other words, PMV can be predicted accurately, but the uncertainty can be increased in actual buildings.

As a result, we found that the temperature and humidity (and air velocity, if possible) commonly measured in real buildings can be used to predict the thermal comfort of occupants using GP. These results represented that RMSE = 0.0183, MBE = -0.00734, and SSE is nearly 0.0003, which shows that the predicted and actual values are very close (Table 4).

Table 4: Predicted PMV in an actual building

<table>
<thead>
<tr>
<th>Performance</th>
<th>RMSE</th>
<th>MBE</th>
<th>SSE$_{\text{mean}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.0183</td>
<td>-0.0734</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Conclusions

We identified the influences of the environmental factors on the thermal comfort of an occupant in a room and employed them to develop a data driven model using the GPR. Hence, the GPR approach used in this study will help easily and efficiently recognize the thermal comfort of occupants by using the data monitored in the building. In the test-bed experimental study, the results of the thermal comfort prediction model using the Gaussian process regression showed that PMV, which is a thermal comfort indicator, can be accurately predicted with relative humidity and ambient temperature variables that are easily measured in real buildings. When the developed model is applied to real buildings, the predicted PMVs are very close to the measured PMV values as well. However, in real building, we found that the 95% confidence interval appears somewhat wider than that obtained in the test bed experiment. In other words, the prediction of the thermal comfort in the real building may increase the associated uncertainty. This is because in real building, the room temperature and relative humidity are affected by many other factors, including the building structure and materials, surrounding environment (e.g., ambient temperature, humidity, and solar radiation intensity), and internal heat gains, occupant behaviour, and so on.

Previous data-driven models PMV prediction with measured environmental data do not provide such probabilistic distribution and uncertainty information. Thus, it is difficult to control thermal comfort considering an uncertainty that is not satisfied by adjusting the boundary PMV values. However, the GP-based PMV prediction model developed in this study estimates accurate PMV predictions, but also provides uncertainty information of the predicted values. This information can be used for maintaining reliable and comfortable space with environmental control devices in real situations.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2018R1A2A2A05023124).

References


