Scientific Research Publishing

Personalized Thermal Comfort Model with Decision Tree

Yuze Jiang

Cushing Academy, Ashburnham, MA, USA
Email: 2198589712@qq.com

Abstract

Thermal comfort is the expression of people's satisfaction with the indoor temperature and is related to people's working efficiency and health. In this way, it is necessary to construct a suitable environment for the user. However, even if adaptive thermal comfort has been developing rapidly for the past decades, most of the models are still developed based on simple statistical analysis such as regression models, which may not capture the complex relations between thermal comfort and the indoor thermal environment as well as differences between individual characteristics. Hence, in order to improve the accuracy of the adaptive thermal comfort model, this paper proposes a decision-tree-based thermal comfort model developed with the subset of the RP884 dataset. Then, a comfort-based HVAC controller was developed with the thermal sensation prediction results with the trained model above. As a result, the proposed controller indeed improves occupant's thermal comfort model.

Keywords

Thermal Comfort, Decision Trees, HVAC

1. Introduction

Thermal comfort is the expression of people's satisfaction with the indoor temperature and is related to people's working efficiency and health. In this way, it is necessary to construct a suitable environment for the user. Thermal comfort is related to six factors, including indoor air temperature, indoor mean radiant temperature, airspeed, relative humidity, clothing insulation, and metabolic rate. Most of the existing HVAC systems are being operated with fixed schedules, pre-assumed occupant activities, pre-assumed clothing, and the assumption of

the maximum number of occupants during the peak period. However, the fixed schedule is not good enough since they cannot reflect the dynamic situations and may result in inefficiency of HVAC operations. Therefore, the usage of responsive HVAC operations with real-time occupant related information is necessary and it is able to improve building performances, including energy savings and occupant comfort level. Lu *et al.* proposed the dynamic HVAC operations based on the number of occupants by occupant recognition with YOLO and taking both energy efficiency and thermal comfort into account and conducted an energy simulation to show the energy savings with the real-time occupant estimation system [1]. In addition, Bing Dong set up a large-scale sensor network to implement a nonlinear model predictive control [2]. Moreover, Zhao has developed and demonstrated the concept of Design-Build-Operate Energy Information Modeling (DBO-EIM) infrastructure, which involves occupant-oriented predictive building control for reducing energy consumption and maintaining occupant thermal comfort [3]. However, most of studies on dynamic HVAC operations apply static thermal comfort model into the control logics, which cannot represent dynamic conditions in the real built environment.

1.1. Static Thermal Comfort vs Adaptive Thermal Comfort

In the course of thermal comfort research, thermal sensation scale is used for quantifying occupants' thermal sensation. The thermal sensation scale goes from "−3" to "3". "−3" means the person feels very cold, and "3" means the complete opposite. In addition, thermal comfort models can be classified into mainly 2 categories. One is a static thermal comfort model and the other is an adaptive thermal comfort model.

In static thermal comfort model, predicted mean vote (PMV) is the most well-known model to predict and evaluate occupant thermal sensation in the indoor environment. The predicted mean vote was developed by Povl Ole Fanger at Kansas State University and the Technical University of Denmark as an empirical fit to the human sensation of thermal comfort [4]. The prediction of thermal comfort involves Fanger's Equation:

$$
\begin{aligned}
PWV = \left[0.303\mathrm{e}^{-0.036M} + 0.028 \right] &\left\{ (M-W) - 3.96E^{-8} f_{cl} \left[(t_{cl} + 273)^4 \right. \right. \\
&\left. - (t_r + 273)^4 \right] - f_{cl} h_c (t_{cl} - t_a) - 3.05 \left[5.73 - 0.007(M-W) - p_a \right] \\
&\left. - 0.42 \left[(M-W) - 58.15 \right] - 0.0173 M (5.87 - p_a) - 0.0014 M (34 - t_a) \right\}
\end{aligned}
\tag{1}
$$

where

PMV is predicted mean vote,

M is the metabolic rate, [met],

t_a is the average indoor air temperature, [˚C].

In addition, PPD describes the percentage of dissatisfaction:

$$
PPD = 100 - 95\mathrm{e}^{\left[-\left(0.3353 PMV^4 + 0.2179 PMV^2 \right) \right]}
\tag{2}
$$

1.2. Adaptive Thermal Comfort vs Adaptive Thermal Comfort

On the contrary, adaptive thermal comfort adapts to the differences between people, such as gender, age, and nationality. For instance, a bio-sensing mechanical system controls were developed based on the experiments regarding the relations between the individual thermal comfort and the skin temperature of different parts of the human body [5]. Moreover, one of the milestones in adaptive thermal comfort research is the development of the database called ASHRAE RP884 database [6]. The introduction to RP884 helps to examine the thermal sensation, acceptability, and preference.

Human-in-the-loop (HITL) is a model that requires human interaction. This can be simply interpreted by this project about thermal comfort. The Thermal Comfort is human's feeling about an environment, which means there has to be someone responding every change in the environment. Human-in-the-loop is efficient in training a model, and it is able to attempt a possible outcome of a specific experiment. HITL is the only way that can reduce the human error in an experiment. Human can judge the process accurately compare to mechanic studies, so that it is necessary to use Human-in-the-loop. Recently, since human-machine interaction has been developed rapidly, researchers tend to focus on developing occupant-centric and intelligent system so as to shape the indoor temperature and humidity and provide comfortable environment. Among different occupant-centric control strategies, rule-based control is one of the simplest yet effective control logics. Rule-based control is a way of using domain knowledge to design the rules so that the system could automatically output the desired actions. The control can be easily used in the investigation of controlling human thermal comfort automatically by adjusting the thermal environment with feedbacks from occupants.

Moreover, many researchers in HVAC control has investigated much on reinforcement learning control (RL) recently. RL problems involve learning what to do—how to map situations to actions—so as to maximize a numerical reward signal [7]. Similar to rule-based control, reinforcement learning control is also a closed-loop control while it includes the learning process instead of existing rules based on domain knowledge. In occupant-centric control system with reinforcement learning, agent must be able to sense the state of the occupant sensation to some extent and must be able to take actions to affect the state. In addition, the agent must be inspired to reach the optimal occupant thermal comfort state by getting the largest rewards.

1.3. Research Contributions

Even if adaptive thermal comfort has been developing rapidly for the past decades, most of the models are still developed based on simple statistical analysis such as regression models, which may not capture the complex relations between thermal comfort and the indoor thermal environment as well as differences between individual characteristics. Hence, in order to improve the accu-

racy of the adaptive thermal comfort model, this paper proposes a decision-tree-based thermal comfort model developed with the subset of the RP884 dataset. Then, a comfort-based HVAC controller was developed with the thermal sensation prediction results with the trained model above.

2. Methodology

2.1. Development of Thermal Comfort Model

In order to do the thermal comfort model, a cleaned dataset with responses from more than 5000 people is used in order to predict the model. The dataset is separated into training data and testing data to see if the model fits various people's thermal comfort about a certain temperature. The control that used in this model is rule-based control, and that's commonly used in the control system. Operators can store the model and allow it to be used. The control is related to the method that people can experience the most suitable thermal comfort without the need to control anything themselves. At the end of the project, the goal is to control over temperature set-point automatically.

2.1.1. Data Description

The original dataset of RP884 dataset [6] has numerous attributes such as gender, birthdate, time, and their subject number, etc. and over 22,000 sets of data. However, after feature selection and preprocessing, the dataset has been trimmed with the most relative attributes and recorded over 5000 responses. Table 1 shows the selected attributes of the thermal comfort model and the label used in the development of thermal comfort model. In addition, the labels are recorded based on the thermal sensation scale, as shown in Table 2.

2.1.2. Decision Tree Thermal Comfort Model

In the decision tree thermal-comfort model, tools such as scikit-learn, pandas and numpy have been used in order to train a suitable model. Scikit-learn is mainly for data mining and data analyzing and the method is built on numpy [8]. Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Numpy is the fundamental package for scientific computing with Python. It contains a powerful N-dimensional array object, sophisticated (broadcasting) functions, tools for integrating C/C++ and Fortran code, useful linear algebra, Fourier transform, and random number capabilities. These packages are fundamental programming tools in the construction of such statistical thermal comfort model. Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features [8].

Decision Trees has some advantages than other methods for classification. It's easy to understand and interpret as the trees are visualized. The process is also less complicated with less preparation. The decision tree allows it to be possible

Table 1. The feature used in the development of thermal comfort model.

Variable	Explanation
ash	ASHRAE Thermal Sensation Scale
met	average metabolic rate of subject
insul	clothing plus upholstery insulation
taav	average of three heights' air temperature
rh	relative humidity
dayav_ta	average outdoor air temperature
dayav_rh	outdoor air relative humidity

Table 2. Thermal sensation scale.

−3	−2	−1	0	1	2	3
Cold	Cool	Slightly Cool	Neutral	Slightly Warm	Warm	Hot

to account for the reliability of the model with static test, and the decision trees always perform well. Nevertheless, disadvantages always campaign advantages to exist. It's possible to overfit the decision tree so that the method does not run well to generate data. The decision tree is also unstable because of small variations in the data, so that a different tree can be possibly generated. There are also specific concepts hard to the decision trees to express. These are all the problems of decision trees. However, it's the best suited method here in the investigation.

In order to learn decision tree, different criterions have been utilized to split the trees such as gini and entropy. Gini is commonly used in the classification and regression tree (CART). It measures how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. On the other hand, information gain is also commonly used to find out the best features for splitting. Information gain measures how much "information" a feature gives us about the class, which consists of entropy, as shown below:

$$I(Y, X_i) = H(Y) - H(Y|X_i) \qquad (3)$$

where

$I(Y, X_i)$: Information gain between label Y and the instance X_i.

$H(Y)$: Entropy of Y.

$H(Y|X_i)$: Conditional entropy of Y given X_i.

Decision tree algorithm is introduced to create a trained model for thermal comfort so that the program is able to predict the thermal comfort of a person by knowing the temperature, humidity, clothing insulation, etc. In this way, the temperature could be automatically changed to reach people's thermal comfort. Unlike regression model, the thermal comfort model is a classification model where thermal sensation is the discrete classes.

Model Selection contains a process of cross-validation in order to find out the

best model. Hyper-parameters are parameters that are not directly learnt within estimators. In scikit-learn they are passed as arguments to the constructor of the estimator classes. In our model, after 5-fold cross-validation, parameters such as max depth, criterion, are tuned so that the best model can be performed on the test data.

2.2. Comfort-Based Temperature Control

The rule-based control is to allow the program to control a thing depends on a rule provided, which is the comfort-based temperature rule here. With the rule, the temperature is going to keep increasing and decreasing until it reaches a comfortable temperature. The rule-based control is necessary as it's a huge part in the automation of changing temperature to reach thermal comfort.

The following Figure 1 describes the rule-based temperature set-point control logic, starting from the thermal sensation prediction from the trained thermal comfort model. In detail, the thermal comfort prediction first tests if the thermal comfort prediction equals 0, which is the comfortable state. If so, the temperature now is the thermal comfort temperature the operator wants. If the thermal comfort prediction is not equal 0, the temperature then goes into the controller and then goes into the actuator to change the value of the temperature. The next step is to test if the thermal comfort prediction of the new temperature is equal to 0. The same step happened again and again until a thermal comfort temperature is found or exited the loop after 1000 iterations.

3. Result Analysis

3.1. Evaluation of Thermal Comfort Model

- parameter tuning

As shown in Table 3, the increasing value of max-depth generally increases the precision of the model and the default gini works better here than entropy. In addition, the best model has max depth of 9, gini criterion whose cross-validation score of 0.43.

The following Table 4 shows the F1 score, recall and precision of the best model. Recall is that the true positive divided by the total actual positive. Precision is that the true positive divided by the total predicted positive. It talks about how precise your model is out of those predicted positive. F1 score is harmonic mean of recall and precision. It is used for seeking balance between recall and precision. The F1, recall and precision of this best model is 0.41, 0.49 and 0.43, respectively.

- confusion matrix

Confusion matrix is a table that is used to describe the performance of a classification model on a set of test data for which the true values are known. From this confusion matrix of thermal sensation in Figure 2, it is found that majority of correct classifications comes from the predictions with the true label of 0, which is reasonable since only a few occupants complain about the indoor thermal environment.
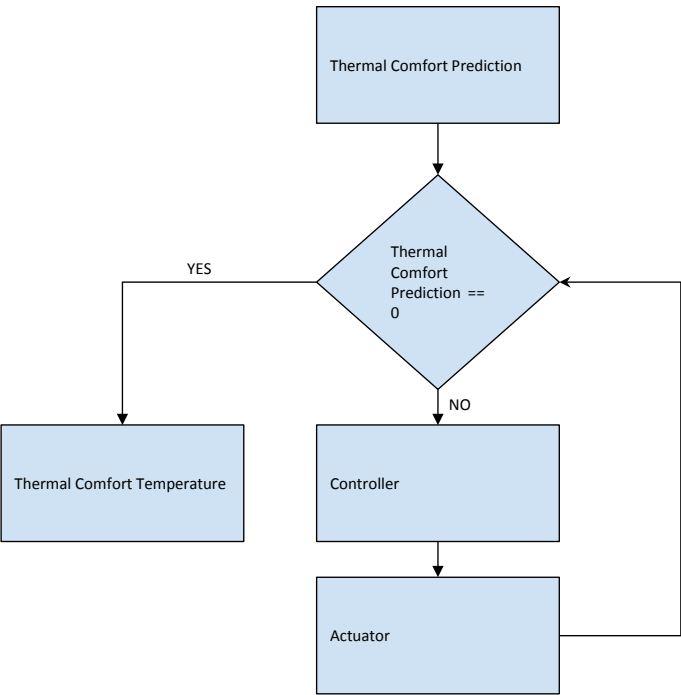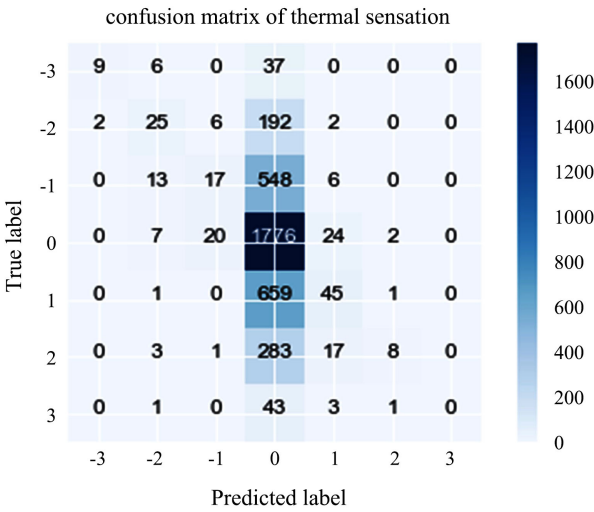
Figure 1. Rule-based control diagram.

Figure 2. Confusion matrix of thermal sensation prediction.

Table 3. Result of model selection.

max_depth	criterion	cross-validation score	F1
5	entropy	0.35	0.36
9	entropy	0.35	0.37
5	gini	0.39	0.36
9	gini	0.43	0.41
7	gini	0.41	0.38
7	entropy	0.36	0.37

Table 4. Performance of the best model.

F1	recall	precision
0.41	0.49	0.43

As in the office environment, occupant thermal comfort is largely dependent on the changes of indoor air temperature since other factors like humidity, clothing insulation are usually steady. Therefore, it makes sense to select the temperature set-point as the controlled variable to adjust the indoor environment to adapt thermal comfort requirements.

3.2. Demonstration of Rule-Based Temperature Control System

As discussed above, the inputs of the prediction model include the following variables: clothing insulation, temperature set-point, indoor relative humidity, outdoor air temperature, outdoor relative humidity and metabolic rate. Moreover, it is assumed the all variables except the indoor air temperature setpoint are fixed. Moreover, in order to show the control sequences with the proposed rule-based controller, the changes of temperature set-point with a starting state is shown below (Table 5).

4. Discussion

Beyond the work conducted in the paper, it is meaningful to explore in the following aspects to improve the controller's performances:

- The goal of ensemble methods is to combine the predictions of several base estimators. In that way, compared to the single model used in the paper, ensemble methods is more adapted to real life environment since the ensemble model improves the accuracy of predictions.
- Joint control of temperature and humidity set-point is needed because humidity does change due to a lot of factors. For example, people need to breathe that increases vapor in air. In this way, temperature and humidity should be controlled at the same time. Using trained thermal comfort model, we are able to test if the specific temperature and humidity are suitable for people feeling comfortable.
- Last but not least, compare to rule-based control, reinforcement learning control is smarter and more advanced. It is able to adapt more situations than rule-based control.

5. Conclusions

This paper aims to propose and evaluate a comfort-based temperature set-point controller with a decision-tree based thermal comfort model. Based on the result and discussions, the following conclusions and future work can be drawn:

- The model performed well in improving people's thermal comfort. However, the precision is relatively low. It is expected that with more variables are imported into the model, the precision would grow higher than now.

Table 5. Demonstration of the rule-based temperature set-point control sequence.

sequence ID	clothing insulation [clo]	indoor temperature set-point [C]	indoor air relative humidity [%]	outdoor air temperature [C]	outdoor relative humidity [%]	metabolic rate [met]
1 (starting state)	0.784	15	50	27.55	81.55	1
2	0.784	16	50	27.55	81.55	1
3	0.784	17	50	27.55	81.55	1
4 (end state)	0.784	18	50	27.55	81.55	1

- With rule-based temperature set-point control, the system can adjust temperature automatically with the input of thermal comfort model prediction.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Lu, S., Hameen C.E. and Aziz, A. (2018) Dynamic HVAC Operations with Real-Time Vision-Based Occupant Recognition System. 2018 *ASHRAE Winter Conference*, 20-24 January 2018, Chicago.

[2] Dong, B. (2010) Integrated Building Heating, Cooling and Ventilation Control. Ph.D. Thesis, Carnegie Mellon University, 1-174.

[3] Zhao, J. (2015) Design-Build-Operate Energy Information Modeling for Occupant-Oriented Predictive Building Control.

[4] Fanger, P.O. (1970) Thermal Comfort. Analysis and Applications in Environmental Engineering.

[5] Choi, J.H. (2010) CoBi: Bio-Sensing Building Mechanical System Controls for Sustainably Enhancing Individual Thermal Comfort.

[6] De Dear, R.J., Brager, G.S., Reardon, J. and Nicol, F. (1998) Developing an Adaptive Model of Thermal Comfort and Preference/Discussion. *ASHRAE Transactions*, **104**, 145.

[7] Sutton, R.S. and Barto, A.G. (2018) Reinforcement Learning: An Introduction. MIT Press, Boston.

[8] Pedregosa, *et al.* (2011) Scikit-Learn: Machine Learning in Python. *JMLR*, **12**, 2825-2830.

Notation List

PMV: predicted mean vote

M: metabolic rate [met]

t_a: average indoor air temperature, [˚C]

PPD: percentage of dissatisfaction

$I(Y, X_i)$: Information gain between label Y and the instance X_i

$H(Y)$: Entropy of Y

$H(Y|X_i)$: Conditional entropy of Y given X_i