Electrical Power and Energy Systems 103 (2018) 634–643

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Improved thermal comfort modeling for smart buildings: A data analytics study

Wei Zhang[a], Fang Liu[a,*], Rui Fan[b]

[a] School of Computer Science and Engineering, Nanyang Technological University, Singapore
[b] School of Information Science and Technology, ShanghaiTech University, Shanghai, China

ARTICLE INFO

Keywords:
Thermal comfort
Machine learning
Data analytics
Smart buildings
Smart city
Smart grid

ABSTRACT

Thermal comfort is a key consideration in the design and modeling of buildings and is one of the main steps to achieving smart building control and operation. Existing solutions model thermal comfort based on factors such as indoor temperature. However, these factors are not directly controllable by building operations, and instead are a by-product of complex interactions between controllable parameters such as air conditioning setpoint and other environmental conditions. In this paper, we use machine learning (ML) to bridge the gap between controllable building parameters and thermal comfort, by conducting an extensive study on the efficacy of different ML techniques for modeling comfort levels. We show that neural networks are especially effective, and achieve 98.7% accuracy on average. We also show these networks can lead to linear models where thermal comfort score scales linearly with the HVAC setpoint, and that the linear models can be used to quickly and accurately find the optimal setpoint for the desired comfort level.

1. Introduction

Thermal comfort modeling is a crucial component in the process of building design, operation and optimization. Energy consumption by buildings accounts for 40% of energy and 60% of electricity usage worldwide. On average, over 50% of a building's energy is used by the heating, ventilation and air conditioning (HVAC) system [1,2], while in areas such as Australia and the Middle East the figure can be as high as 70% [3]. The primary product of an HVAC system is the thermal comfort. People today spend over 90% of their time in buildings [4], and poor comfort in buildings increases the chances of sick building syndrome, absenteeism and cognitive degradation [5]. Thus, it is important to create a healthy and comfortable indoor space, while at the same time minimizing building energy use. A key step towards this goal is creating accurate models of thermal comfort.

In the past decades, thermal comfort modeling has received much research attention. The most popular model is the predicted mean vote (PMV) model proposed by Fanger et al. [6] and adopted in the ASHRAE Standard 55 [7]. PMV models thermal comfort using six factors, including four environmental factors (indoor temperature, indoor humidity, mean radiant temperature (MRT) and air velocity) and two vital ones (metabolic rate and clothing insulation).

Although the PMV model works fine for evaluating thermal comfort, it is not readily applicable to smart buildings. In a typical smart building, the building management system (BMS) predicts thermal comfort levels based on different controllable building settings, before deploying specific settings. However, while the PMV model establishes a comfort score based on the PMV factors, it does not capture the relationship between controllable building settings and the comfort score. In particular, the three indoor PMV factors, indoor temperature, indoor humidity and MRT, cannot be directly controlled by the BMS, and instead result from complex interactions between controllable building parameters such as HVAC settings, weather conditions and other factors. Thus, to improve thermal comfort for smart buildings, controllable building factors should be used for modeling in place of non-controllable ones.

In this paper, we aim to use machine learning (ML) [8–10] to capture the impact of controllable HVAC operations on thermal comfort. The basic idea is first to use different ML algorithms to model the relationship between controllable parameters and the indoor PMV factors, given the current date, time and weather information. Then, the PMV factors predicted by the ML models for different HVAC settings are fed into the BMS for decision making. To validate the soundness of this approach, we performed extensive data analytics using a variety of ML models. The key results we obtained are as follows:

1. Nonlinear ML algorithms, including support vector machine regression (SVR) with nonlinear RBF kernels and neural networks

* Corresponding author.
E-mail addresses: wei.zhang@ieee.org (W. Zhang), fliu@ieee.org (F. Liu), fanrui@shanghaitech.edu.cn (R. Fan).

https://doi.org/10.1016/j.ijepes.2018.06.026
Received 27 February 2018; Received in revised form 27 April 2018; Accepted 7 June 2018

(NN) perform on average 66% better than linear methods, including linear regression (LR) and SVR with linear kernels. This indicates a complex relationship between the data features and PMV factors.
2. The performance of NNs depends on the network configuration. We found an optimal configuration with 7.2% higher PMV modeling accuracy compared with the default configuration. Furthermore, the optimal configuration can be trained very quickly, in only a few seconds.
3. The ML-based solution using NNs achieves high modeling accuracy, with an average error of only 1.3% in predicted PMV comfort levels. Additionally, we show using the NN models that the PMV score is linearly related to the HVAC setpoint, given knowledge of other non-controllable factors such as date, time and weather conditions. Using the linear model, we can efficiently find the optimal, least energy-intensive HVAC setpoint achieving a desired PMV comfort level.

The rest of this paper is organized as follows. We review related works on comfort modeling in Section 2. We describe our dataset and present the model architecture in Section 3. In Section 4, different ML algorithms are used to model the relationship between the dataset and the PMV factors. In Section 5, we compare the models and show that NNs achieve the best performance. We also show how the NN model can be used for smart building control. Finally, Section 6 concludes the paper and suggests future works.

2. Related works

Thermal comfort modeling has been studied extensively in the past decades. The most well-known thermal comfort model is the PMV model proposed by Fanger et al. [6] and adopted in the ASHRAE Standard 55 [7]. PMV assumes that thermal comfort is determined by six factors, including four environmental factors (indoor temperature, MRT, indoor humidity and air velocity) and two vital factors (metabolic rate and clothing insulation). Based on these factors, a set of equations are given to derive a thermal comfort score ranging from −3 to 3. The seven integers within this range can be interpreted in ascending order as indicating cold, cool, slightly cool, neutral, slightly warm, warm and hot, respectively. Based on the PMV score, a thermal comfort index called predicted percentage of dissatisfied (PPD) can also be computed. The PPD value ranges from 0% to 100%, where small values indicate great comfort. ASHRAE Standard 55 recommends that PMV and PPD should be within $\pm$ 1.0 and $\leq$ 20%, respectively, to meet basic occupant thermal comfort needs.

In addition to the PMV model, ML offers another way to model thermal comfort. Megri et al. in [11] applied the ϵ-SVR algorithm to the six PMV factors to predict the PMV comfort score. Using 793 data samples for training and 18 samples for testing, they achieved a modeling accuracy of up to 99%. Atthajariyakul and Leephakpreeda [12] also modeled PMV scores, but used a NN model with two hidden layers. Their results showed very high modeling accuracy.

The main difference between the past works and ours is that earlier model was parameterized by the PMV factors, whose values are not directly controllable. To be of use for a smart building BMS, it is crucial that a thermal comfort model is based on parameters that the BMS can directly control [13,14]. In this paper, we first model the relationship between the controllable parameters and the PMV factors and then use the predicted factors to compute comfort scores for BMS control and deployment.

3. Dataset and system architecture

In this section, we first describe the dataset we use for training our models, before describing the model architecture.

Table 1
Valid data range and type.

Feature	Min	Max	Type
HVAC1 On/Off	0	1	bool
HVAC2 On/Off	0	1	bool
HVAC1 Setpoint	15	50	float
HVAC2 Setpoint	15	50	float
Outdoor Temperature	15	50	float
Outdoor Humidity	10	100	float
Outdoor Irradiance	−50	1,400	float
Outdoor Illuminace	0	200	float
Rain	0	1	bool
Month	1	12	integer
Weekday	1	7	integer
Hour	1	24	integer
Minute	1	60	integer
Indoor Temperature	15	45	float
Indoor Humidity	10	100	float
Indoor MRT	15	50	float

3.1. Dataset

The data resolution is one minute, meaning that a data sample was recorded every minute. Each data sample has four major feature sets, including the date and time (datetime), weather conditions, HVAC settings and indoor data. The first three sets of data are used as the dependent features of the ML algorithms, while the indoor data values are the target features the algorithms try to learn. Details of each feature set are shown in Table 1, and described below.

3.1.1. Datetime data

Each data sample is associated with a timestamp including the date and time. Since ML algorithms typically deal with numerical values rather than a datetime string, the timestamp is broken into four numbers, including month, weekday, hour and minute. Weekday is a value ranging from 1 to 7 representing the days of the week.

3.1.2. Weather data

The weather data has five features, including the outdoor temperature in Celsius (°C), the percentage of outdoor humidity, outdoor irradiance in the Watt per square meter (W/m^2), outdoor illuminance in lumens per square meter (Lux), and the rain status, either true or false.

3.1.3. HVAC data

The testbed office room has two fan coil units. The status of each unit, either on or off, was monitored and denoted using a Boolean value. Also, the setpoint temperature of each unit in degrees Celsius was recorded as a floating point number. In total, each HVAC datapoint has four values, HVAC1 On/Off, HVAC2 On/Off, HVAC1 Setpoint and HVAC2 Setpoint. The four HVAC features are the only controllable features in our system.

3.1.4. Indoor data

Three indoor features are recorded in the dataset. The first two are the indoor temperature in Celsius and the percentage of indoor humidity. The last feature is the MRT, measured in degrees Celsius. MRT models the heat exchange between an indoor object like the human body and the indoor environment. Past studies have shown that MRT is one of the most critical factors for thermal comfort [6,15].

3.1.5. Data preprocessing

The collected data was often noisy due to the problems such as faulty sensors or software bugs. The dataset was cleaned to ensure the recorded data was valid. The valid data range of each feature is shown in Table 1.