Building and Environment 118 (2017) 323–343

Contents lists available at ScienceDirect

Building and Environment

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

A Bayesian approach for probabilistic classification and inference of occupant thermal preferences in office buildings

Seungjae Lee [a, b], Ilias Bilionis [c], Panagiota Karava [a, b, *], Athanasios Tzempelikos [a, b]

[a] School of Civil Engineering, Purdue University, 550 Stadium Mall Dr., West Lafayette, IN 47907, USA
[b] Center for High Performance Buildings, Ray W. Herrick Laboratories, Purdue University, 140 S. Martin Jischke Dr., West Lafayette, IN 47907, USA
[c] School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 12 January 2017
Received in revised form
21 February 2017
Accepted 4 March 2017
Available online 14 March 2017

Keywords:
Bayesian modeling
Clustering
Inference
Personalized environments
Thermal comfort
Thermal preference

ABSTRACT

This paper presents a new data-driven method for learning personalized thermal preference profiles, by formulating a combined classification and inference problem, without developing different models for each occupant. Different from existing approaches, we developed a generalized thermal preference model in which our main hypothesis, “Different people prefer different thermal conditions”, is explicitly encoded. The approach is fully Bayesian, and it is based on the premise that the thermal preference is mainly governed by (i) an overall thermal stress, represented using physical process equations with relatively few parameters along with prior knowledge of the parameters, and (ii) the personal thermal preference characteristic, which is modeled as a hidden random variable. The concept of clustering occupants based on this hidden variable, i.e., similar thermal preference characteristic, is introduced. The results, based on a dataset collected from a typical office building population, show clear evidence of the existence of multi-clusters; in particular, the 5-cluster model performed best compared to 2, 3 and higher cluster models using the studied dataset. Subsequently, the thermal preference of a new occupant in the dataset is inferred by using a mixture of the general sub-models for each cluster. The results show that the method developed in this study provides accurate predictions for personalized thermal preference profiles and it is efficient as it only requires a relatively small dataset collected from each occupant. The approach presented in this paper is a significant step towards personalized environments in office buildings using real-time feedback from occupants.

1. Introduction

Operation of heating and cooling systems in office buildings has been standardized and automated based on the use of “widely acceptable” thermal comfort metrics and simple heuristic rules. However, field studies have shown that individual occupants prefer different thermal conditions and such metrics could not accurately predict thermal preference of individual occupants [1–4]. As a result, typical thermal control systems cannot achieve high levels of occupant satisfaction. Moreover, because of the conservative control settings designed for “widely acceptable” conditions, there is a high probability of energy waste [5,6]. Researchers have recognized these issues and suggested solutions that incorporate building occupants in sensing and control frameworks and tune systems based on personal preferences to achieve customized indoor environments [7–12]. The method of learning individual occupant preferences is essential and determines the effectiveness of this solution, since system control is based on the learning outcomes.

Recent studies have attempted to explore personalized environments following different approaches. For example, Murakami et al. [7] developed an automatic control logic which maintains a balance between occupants' needs and energy consumption. The logic was implemented in an office space and resulted in 20% of cooling energy savings without increasing the percentage of occupant dissatisfaction. Feldmeier and Paradiso [8] developed a user-centric distributed control system which learns individual occupants' thermal preferences using linear discriminant analysis. The authors tested the system in real offices and reported 24% savings in HVAC energy consumption (compared to standard HVAC control) and improved thermal comfort. Erickson and Cerpa [9] introduced a time-varying variable in the predicted mean vote

* Corresponding author. School of Civil Engineering, Purdue University, 550 Stadium Mall Dr., West Lafayette, IN 47907, USA.
E-mail address: pkarava@purdue.edu (P. Karava).

http://dx.doi.org/10.1016/j.buildenv.2017.03.009
0360-1323/

(PMV) model, which corrected the predictions made by the original model to match occupants' feedback. The authors controlled a space based on this model and reported that occupants were fully satisfied with the thermal conditions while 10.1% energy savings were achieved. Gao and Keshav [10,11] proposed a method to map the PMV value to occupant feedback using a personalized simple regression model. The authors developed a controller by integrating the method with clothing level estimation, occupancy detection and model predictive techniques, and implemented the controller in an office room, with resulting energy savings up to 60%. Jazizadeh et al. [12] developed an HVAC control framework that learns occupant comfort profiles using a fuzzy model based on user feedback. The authors implemented the framework in a real building and showed that 39% of HVAC energy was reduced while occupant comfort was improved.

1.1. Current approaches for learning occupant thermal preferences

In previous research, learning is associated with adapting (estimating) parameters in a model (or a control logic) based on data collected from an individual occupant, so that the model can predict the occupant's thermal comfort status, typically expressed as preference, satisfaction, or sensation under certain conditions [7–17]. However, developing a reliable model requires data with sufficient quality and quantity that is difficult to collect from individual occupants in real buildings, because occupants should not be exposed to potentially uncomfortable conditions for a long time. To address this problem, previous studies have used two strategies, described below.

The first strategy is based on long-term data collection, and consequent use of the data to adapt the personalized model. For example, Daum et al. [14] proposed a learning method based on multinomial logistic regression and showed that personalized profiles were needed to predict occupant's thermal sensation properly. In this study, the authors used data collected from 28 different occupants during a period of three years. The total number of data points was 6851. The authors also investigated how a personal comfort profile evolved as new information was collected and used. The results showed that the method required more than 90 data points to develop converged personalized profiles, which might take more than three months under normal office conditions. However, the long-term data collection often translates into long times required for developing functionally advanced controls, depending on the learning outcomes, while data quality is not guaranteed. For example, if data is collected in a space with consistent thermal conditions, the data may not represent complete occupant thermal preference patterns, and model parameters might not always be estimated properly. For instance, the personalized profiles in Ref. [14] did not evolve significantly with data collected in January since the room temperature was kept within a comfortable range and the occupant answered "comfortable" during most of the time in that month.

The second strategy is based on the simplification of the model structure so that it can be developed with fewer data. Although studies have been also conducted using complex models [10,11,13,17], they did not demonstrate if model predictions were reliable and how potential overfitting problems (due to small-sized datasets) were addressed. Most of these studies used models having only one input variable: air temperature [8,12,14,15]. The rationale behind this method is the hypothesis that a prediction made with the air temperature is not inferior to using a complex model in typical buildings, which was first reported by Humphreys and Nicol [18]. Even though other environmental factors may highly co-vary with air temperature or remain constant in typical buildings, since simplified models do not consider the metabolic rate and thermal insulation level of clothing, their prediction performance may be limited. To resolve this problem, authors in Refs. [12,14,15] suggested updating the models continuously by discarding old data so that the models could reflect adaptive behaviors and physiological adaptation. Although these models may adequately reflect occupant adaptation, that is only possible after occupants express discomfort. In other words, this approach might not always be sufficiently predictive, and the models do not take into account the effect of adaptation before occupant feedback. Therefore, it is an inherent limitation that these models might not predict adaptation and, if a space is controlled based on this approach, occupants may be continuously exposed to uncomfortable conditions. Although the approach is straightforward and practical, with regards to the final goal, it cannot deliver personalized indoor environmental conditions for occupant satisfaction, which might reflect that the approach is self-limited. Moreover, if other environmental factors (i.e., MRT, air velocity, humidity) do not highly co-vary with air temperature or do not remain constant, the simple model may not provide reliable predictions. For example, in perimeter zones, occupants are affected by both solar and longwave radiation which may vary significantly depending on sky conditions and solar exposure of the person [19–21]. Also, in some cases, the HVAC system controls not only the air temperature but also other parameters in order to create comfortable thermal environments (e.g., radiant heating/cooling systems or local systems exploiting the effect of increased air velocity) [22]. Considering these cases, air temperature-based models may be too simple to accurately predict occupant thermal preferences.

1.2. Research contributions

This study presents a novel Bayesian modeling approach for learning individual occupants' thermal preferences, with improved efficiency and accuracy, even if the model structure is complex and the amount of data collected from each occupant is relatively limited. The approach includes (i) developing a general model structure to explain the thermal preferences for a population of occupants (Section 2); verification with a synthetic dataset collected in various office buildings, finding the optimal number of clusters, and results based on a subset of the large ASHRAE RP-884 database [24] (Section 3); (ii) using the general model to infer the personalized thermal preference profile of an individual occupant and prediction results for personalized thermal preference profiles (Section 4). The prediction performance of the overall method along with limitations and recommendations for future work is discussed in Section 5. To enable this approach, we explicitly encode our main hypothesis: "Different people prefer different thermal conditions" in the general model. The rationale behind our modeling choice, i.e. a Bayesian modeling approach, is related to its inherent advantages: it allows encoding and testing our prior knowledge and beliefs about the relationships between the different variables; it can easily account for hidden (unobserved) variables; and it can seamlessly combine data from heterogeneous sources as they become available [23]. Fig. 1 shows the overall process for learning the thermal preference of an individual occupant.

2. The general model: discovering clusters of occupants with similar thermal preference characteristics

2.1. Modeling methodology

An important starting point for developing a model is drawing relationships between variables. Fig. 2 is a simplified version of the graph representing connections between thermal preference and