# Input Output HMM for Indoor Temperature Prediction in Occupancy Management Under User Preferences 

Kamrul Islam Shahin<br>University of Southern Denmark, Odense, Denmark<br>kish@mmmi.sdu.dk

Anooshmita Das<br>Machine Learning Lead, Liveintent<br>Inc. Copenhagen, Denmark<br>adas@liveintent.com

Mikkel Baun Kjærgaard<br>University of Southern Denmark, Odense, Denmark<br>mbkj@mmmi.sdu.dk


#### Abstract

In this paper, a probabilistic machine learning method is proposed to predict the indoor temperature of an office environment. An IOHMM-based model is developed to represent the office environment under different circumstances of heating sources. One year of time series data is observed and studied to learn the dynamics of the indoor thermal states. The uncertainty associated with the changing aspects of the indoor temperature and its dependence on the outdoor temperature is considered in the model development. The well-known Baum Welch and forward-backward algorithms are adapted to learn the model parameters. Then, the Viterbi algorithm is used to predict the maximum path of hidden states, leading to predicting the most likely future temperatures. A numerical application is presented to demonstrate the model development steps and the training and testing results. Finally, the model's performance is validated using leave-one-out cross-validation, which shows that the model has a prediction accuracy of about $78 \%$.


Keywords: Building occupancy, data monitoring, office action, indoor temperature prediction, HMM

## 1. Introduction

Reaching carbon neutrality goals require that our use of office building real-estate become more energy efficient. Different studies (Rahaman, 2019; Alrazgan, 2011; LeMay, 2009) explore how to optimize the use of space using digital tools both so occupants require less space and to only provide comfort in used spaces to more efficiently spend energy. Rahamen (2019) studies how to sense the used spaces of people to optimize for this and Kahn (2020) explore how to optimize the personal comfort of the space used by occupants. Another interesting concept is free seating where the seating arrangement of occupants are made flexible to lower space use. In particular, when many occupants do not on a regular basis come to office either because they have tasks outside the office or work from home this can
lower the need for real-estate space. A challenge is hot, giving the occupants options to pick a good seating that matches their preferences. Sood [2020] have proposed a system named space match to help occupants with this. This tool and evaluation highlight good opportunities for such tools to help the occupants. However, the space match only provides a match based on what happens previously. In this paper we would like to propose a system that also considers the future development of the indoor environment to provide better matches. Previous work has considered prediction of the indoor environment for control systems (Perić, 2021; Tariq, 2019; Arendt, 2018; Peng, 2016; Ellis, 2003). However, previous work has not considered the design of such prediction algorithms for free-seating systems.

This paper presents a thermal predictor based on the input-output hidden Markov model (IOHMM). The IOHMM structure was proposed by Bengio (1995), and the authors developed a training procedure based on the expectation maximization (EM) algorithm. The model is similar to the Hidden Markov Model, but its advanced form supports recursive processing of input and output events and allows supervised learning models using maximum likelihood estimation. Later, the model was used in various applications (Shahin, 2019; Weber, 2016; Hu J, 2015; Ourston, 2003).

In this paper, the IOHMM architecture is adapted to predict indoor thermal conditions by monitoring sensor data for one year. The training method was designed by using the Baum Welch algorithm (Pathak, 2006) which is a class of EM algorithms and forward-backward algorithms. Afterwards, the model applies the Viterbi algorithm and the Markov characteristic to predict current events and upcoming events, which specifies hidden processes determined by several input conditions. Finally, a preference algorithm matches the user's preferences with the predicted thermal conditions of the indoor environment and suggests desirable seating options to the user.

This paper is organized as follows: Section 2 represents the current state of the art. Then section 3 discusses the model structure and the algorithm
developments. Section 3 describes the data analysis and processing. After that, Section 5 gives a numerical application to an open dataset and finally, Section 6 gives a conclusion.

## 2. Related work

As our collective interpretation of "work" and "the workplace" evolves, office space usage becomes even more complicated. A dynamic trend in office space usage has resulted from new paradigms in working culture. Understanding how spaces serve us, seating preferences, and estimating demand for those spaces allows us to make prompt and informed decisions for better space optimization. Jens (2004) investigate the trade-offs between open and enclosed spaces and how opposing and complementary design aspects influence behavior and occupancy seating preferences. Rahaman $(2020,2019)$ discovered that the lack of preferred seating arrangements could increase workers' perceived stress and impair their focus levels. However, there was no way of knowing how various people's preferred seats could be characterized.

Chafi (2020) conduct similar research in Swedish Municipalities to identify workspace preferences in flexible offices. He also investigates whether employees' workstation selections endorse their activities and align with their preferences. This study emphasized the competition for desirable workspaces using 27 semi-structured interviews and annotations on architectural drawings to obtain data. A variety of factors influenced preferences, including functional, social, emotional, and symbolic considerations. According to the findings, interviewees preferred workstations that were both favorable and functional, while ignoring workstations that were neither desirable nor functional. This was due to the geometry and architecture of the physical space being influenced by various stimuli. The authors concluded that the characterization of underutilized spaces can help improve space configuration. Flexible office layout can be improved by characterizing and replacing undesirable workspaces with more advantageous and appealing alternatives. This method can be used both during the design process and after the relocation to identify and replace undesirable workstations with desirable ones, as well as to alleviate the stress affiliated with finding a suitable workstation in office buildings. Another interesting study by Sood (2020) depicts the implementation and testing of the Space match platform, which was intended to optimize workspace allocation and management. Using a web-based mobile application, this methodology associates' occupants with a listing of available work desks and allows them to provide real-time environmental feedback. Over the
course of a 30-day study, this case study implementation collected 1,182 responses from 25 field-based research participants. The findings reveal that based on their gathered preference data, occupants may be split up into distinct sorts of users, and matching preferences can be generated to develop a recommendation platform using this initial data set. The shortcoming of this investigation is that the representative sample size of participants is insufficient to provide more generalizable categorizations of the different sorts of comfort and the wide range of behavior that occupants can display. In this paper, we offer a method that takes into account the indoor environment's future development in order to deliver better matches. The method also offers the advantage of an open dataset to solve data scarcity issues.

Prediction of the indoor environment for control systems has been studied in the past (Huang, 2020; Alawadi, 2020; Hietaharju, 2018; Sanandaji, 2014). Previous research, on the other hand, has not taken into account the design of such prediction algorithms for free-seating systems. Lu (2019) uses machine learning techniques such as k -nearest neighbor (KNN), random forest (RF), and support vector machine (SVM) to construct a thermal comfort model for three major climate zones utilizing RP884 (open dataset). The statistical thermal comfort model was then used to simulate a tabular Q-learning temperature set-point control system. According to the findings, the statistical thermal comfort model has the best recall of 49.3 percent, which exceeds Predicted Mean Vote's recall of $43 \%$ based on a 7 -point thermal sensation scale. Furthermore, regardless of the initial temperature setpoint, the Q -learning reliant temperature control can reach pleasant temperature ranges for occupants. Ma (2021) applied a Bayesian neural network (BNN) algorithm to build a predictive model for occupant thermal preference using the ASHRAE Global Thermal Comfort Database II. The findings imply that coupling occupants' subjective evaluations and window opening/closing behavior with thermal comfort modelling enhances predictive performance significantly. Fang (2021) used an LSTM-based seq2seq model for accurate indoor temperature forecasting. When compared to Prophet and a seasonal naive model, the LSTM model proved to be significantly more competent and accurate in extremely short-term predicting. To enable multi-zone indoor temperature forecasting with a more generalized model, a cross-series learning technique was used. To measure the uncertainty in model parameters, the Monte-Carlo dropout (MC-dropout) technique was applied.

This under-researched subject of not taking into account the design of such prediction algorithms for free-seating systems inspired the authors to propose a
system that considers the indoor environment to produce better matches for occupant seating. The approach in this research paper is interesting as it uses IOHMM to combine grouping occupants based on their preferences with indoor temperature prediction for short- and longterm temporal resolution. The IOHMM model is better suited for our use-case, and it is evaluated against a baseline HMM model. IOHMM has low-overhead and low-complexity advantages for training purposes as compared to other ML models.

It can also consider pre-decided input settings (time, heating conditions, etc.) to predict the corresponding future events for flexible user preferences. One of the major advantages of IOHMM can be addressed as its capability of handling multiple observations along with the input conditions. It is capable of study several observations (sensor readings: indoor-outdoor temperature, humidity, air pressure, etc.) separately and together (with the dependency between them) which can be extended to the next developments for more efficient predictions (Shahin, 2020).

## 3. Input-Output Hidden Markov Model

The Input-Output Hidden Markov Model is a stochastic model, which is an advanced version of the Hidden Markov Model (HMM). While HMM provides a single model-version for all conditions to predict a mean value, IOHMM is capable of predicting different values for different conditions, e.g., multiple room temperatures can be modelled by a single IOHMM applying input indexes. It can also anticipate multiple outputs corresponding to the same input(s).

IOHMM was previously used to develop a predictive model for time series data (Shahin, 2020). It is introduced in the current contribution to model a different dataset to predict temperature considering different conditions. One of the major advantages of using IOHMM is its training method. For example, it can use multiple data subsets in one training session to learn multiple versions of the model representing the corresponding conditions. On the other hand, the HMM would be trained separately for the same conditions.

### 3.1 Model property

The general properties of an IOHMM are described in detail in (Shahin, 2020). Here, the structure of the model is defined (Figure 1) according to the given data and the problem formulation.


Figure 1. Input Output HMM

- The $Y$ node represents the observation sequence(s) or in this case the indoor temperature measurements converted from continuous to discrete format (see Figure 2) in accordance with the IOHMM properties.


Figure 2. Example: continuous to discrete conversion
Five discrete symbols are used to convert the continuous sequence into discrete format. However, the number of discrete symbols can be less or more than five based on the number of changing events between these symbols. A symbol with no (or not enough) transitions holds emission parameters with poor probability which may leads false predictions.

- The $X$ node is the sequence of hidden states (e.g., representing different scenarios of indoor temperature).
- The $U$ node is for the input conditions.
$-k$ is the sequence length.
- Transition matrix: it represents the transition probabilities $\left(a_{i j}\right)$ from state $i$ to state $j$ for $1 \leq i, j \leq N$; where $N$ is number of hidden states. If the number of hidden states is 4 then the size of this matrix is 4 by 4 square.
- Emission matrix: it represents the emission probability $\left(b_{j l}\right)$ from state $j$ to observable symbol $l 1 \leq l \leq M$; where $M$ is the number of discrete symbols. Let's assume the number of symbols is 5 . So, the size of this matrix is 4 by 5 (number of discrete symbols by number of hidden states).


### 3.2 Model Training

IOHMM uses the Baum Welch algorithm and the Forward-Backward algorithms to be trained and learn its parameters (Rabiner, 1989). The adapted version of the algorithms to IOHMM which originally dedicated to HMM are given below.
The forward backward algorithm:

## Forward part:

Basis: $\alpha\left(X_{1}\right)=P\left(Y_{1} \mid X_{1}\right) P\left(X_{1}\right)$
Recursion: $\alpha\left(X_{k}\right)=$ $\sum_{X_{k-1}=s_{1}}^{s_{N}} \alpha\left(X_{k-1}\right) P\left(X_{k} \mid X_{k-1}, U_{k-1}\right) P\left(Y_{k} \mid X_{k}\right)$, here $\alpha\left(X_{k}\right)$ is the forward auxiliary variable $=P\left(X_{k}, Y_{1: k}\right)$, and k is the length of the observation sequence.

## Backward part:

Basis: $\beta\left(X_{K}=s_{i}\right)=$ all 1 .
Recursion: $\beta\left(X_{k}\right)=$
$\sum_{X_{k+1}=s_{1}}^{S_{N}} \beta\left(X_{k+1}\right) P\left(X_{k+1} \mid X_{k}, U_{k}\right) P\left(Y_{k+1} \mid X_{k+1}\right)$
here $\beta\left(X_{k}\right)$ is the backward auxiliary variable $=$ $P\left(Y_{k+1: K} \mid X_{k}\right)$.
The Baum Welch algorithm:
The Baum-Welch algorithm uses $\alpha\left(X_{k}\right)$ and $\beta\left(X_{k}\right)$ to update the parameters repeatedly in three steps:
Step 01: Initial state probability: $\pi_{i}=\varepsilon_{1}(i, j)$,
where $1 \leq i \leq N$
Step 02: Transition probabilities: $\hat{a}^{p}{ }_{i j}=$ $\frac{\sum_{k=1}^{K-1} \varepsilon_{k}(i, j) \cdot 1_{\boldsymbol{X}_{\boldsymbol{k}}\left(U_{\boldsymbol{k}}=p\right)}}{\sum_{k=1}^{K-1} \omega_{k}(j) \cdot 1_{\boldsymbol{X}_{\boldsymbol{k}}\left(\boldsymbol{U}_{\boldsymbol{k}}=\boldsymbol{p}\right)}}$; where $1_{X_{k}\left(U_{k}=p\right)}=\left\{\begin{array}{l}1 \text { if } X_{k}\left(U_{\boldsymbol{k}}=p\right) \\ 0 \text { others }\end{array}\right.$, $p$ is the number of hidden states
Step 03: Emission probabilities: $\hat{b}_{j k}=$
$\frac{\sum_{k=1}^{K} \omega_{k}(j) .1_{Y_{k}=v_{m}}}{\sum_{k=1}^{K} \omega_{k}(j)}$; where $1_{Y} q_{k}=v_{m}=\left\{\begin{array}{l}1 \text { if } Y^{q}{ }_{k}=v_{m} \\ 0 \text { otherwise }\end{array}\right.$
Here $\omega_{k}(j)$ is the probability of being in state $j$ at time $k$ given the sequences $Y$, and $\varepsilon_{k}(i, j)$ is the probability of being in states $i$ and $j$ at time $k$ and $k+1$ given the sequences $Y$.

### 3.3 Prediction

The given data provide only the observation sequences, but not the (hidden) state sequence. This is a sequence of unknown states or situations producing the temperature which is recorded in a regular time interval. This is why the state sequence has the same length as the observation sequence.

Since the state sequence is hidden or unknown it is needed to be predicted which let the model to predict the future state evolution and the corresponding emissions at the current time given the current observations.

The IOHMM uses the Viterbi algorithm, which is a popular algorithm to predict the maximum path of the hidden states given the observation sequence. The
maximum path characterizes the sequence of states used to predict the future hidden states and their output temperature applying the Markov property. This algorithm is adapted to IOHMM [25].
The adapted Viterbi algorithm:
Basis: $\gamma\left(X_{1}\right)=P\left(X_{1}, Y_{1}\right)$;
Recursion: $\gamma\left(X_{k}\right)=$
$\max _{\left(X_{k-1}\right)} P\left(Y_{k} \mid X_{k}\right) P\left(X_{k} \mid X_{k-1} U_{k-1}\right) \gamma\left(X_{k-1}, U_{k-1}\right) \gamma\left(X_{k-1}\right)$ , here $\gamma\left(X_{k}\right)=\max _{\left(X_{1: k}\right)} P\left(X_{1: k}, Y_{1: k}\right)$.

This algorithm computes the maximum likelihood path as $P\left(X_{1: k} \mid Y_{1: k}, U_{1: k}\right)$. It contains the probability of $P\left(X_{k} \mid Y_{1: k}, U_{1: k}\right)$ which is the current state distribution at time $k$. Then the transition probability (transition matrix) is used to predict the state at time $k+1$ as $P\left(X_{k+1} \mid X_{k}\right)$. Finally, the maximum probability of emitting the temperature at time $k+1$ is calculated using the emission probability (emission matrix) as $P\left(Y_{k+1} \mid X_{k+1}\right)$.

The last two steps are then repeated until the IOHMM reaches a given time at which the prediction should stop. Normally this is the office closing time, e.g., 4 pm.

## 4. Data analysis

The proposed method is designed to predict the indoor thermal condition of an open office space. A total of 365 days of observed data (from 2013-1-1 to 2013-12-31) are analyzed and evaluated in order to make the prediction under similar conditions. This is an open dataset that can be used for a variety of purposes, including the development and validation of occupancyrelated models.

### 4.1 Office environment

The office layout (Figure 3) consists of several semi-enclosed (O2, O4), enclosed (O3), meeting rooms (MR), kitchen (KI) and a multi-person room (O1).


Figure 3. Layout plan of the office floor

A total of eight workstations ( $\mathrm{C} 1-\mathrm{C} 8$ ) is available in the layout, with all desks close to (at least) one window. Staff may have different access facilities (electrical switch control or close to fire-exit, etc.) depending on their desk position. It is also possible that the occupant may feel colder/warmer than others due to the distance to the heater and/or different heating scenarios (Table 1).

Table 1: Example of different heating states

| Season | Heater | Window | No. Of machines |
| :--- | :--- | :--- | :--- |
| Winter | On | Close | Fixed |
| Spring | On/Off | Close | Fixed |
| Autumn | Off/On | Close | Fixed |
| Summer | Off | Open | Fixed |

This is an example of looking into the indoor situation considering only two variables (heater and window) based on different seasons. For the sack of an easy computation and less complexity, all the other heat sources (number of machines, human beings, etc.) are assumed to be unchanged which can be introduced into the model upon to the context. Since more variables can make the model construction more complex, we considered these four combinations which lead to having a 4 state-IOHMM. A feasible number of states should be considered according to with the data amount. Otherwise, the performance of the model may be affected.
Thermal condition: If there is nothing to predict in order to suggest users for their preferable workstation then only a preference database query would be enough. Such as all the information is pre-defined and fixed (as shown by Figure 4).


Figure 4. Layout plan of the office floor [Here $W$ is for window seat, Shrd is for shared room,

## Cont. is for controller access, F.exit is for fire exit. The office time is assumed as 8:00-16:00]

However, it becomes a challenge when a user seeks a thermal conditioned (cooler or warmer) desk not only at the current time but also for the near future. Since the upcoming temperature is unknown and can be influenced by different uncertainties (number of presence in the room, number of machines on/off in the room, environment conditions, etc.) over the time it is very important to predict temperature efficiently. Now the question is how much efficiency is efficient enough? Well in this study, we considered two observations (indoor and outdoor thermal conditions). More observations may increase the prediction efficiency but with the cost of more complexity.

### 4.2 Data processing

The training set (indoor temperature measurements) is prepared by extracting each day of observations as a single sequence. Therefore, the oneyear data became 365 independent sequences (see Figure 5). After that, a pattern between the indoor and outdoor temperatures is acknowledged by analyzing the dynamics of the temperature (see Figure 6).


Figure 5. Indoor data sequences


Figure 6. Outdoor temperature

First, the sequences are classified based on the level of the outdoor readings from low to high (see Table 2). Then, the days having the temperature within these ranges are marked-out in different classes. On this basis, indoor temperature measurements are classified into four training subsets, which represent four types of behavior in the data. An IOHMM with four inputs is developed by using these subsets, where each input represents one of the four temperature classes (from coldest to hottest). Each subset containing as many observation sequences as the number of days.

Table 2. Data classification

| Id | Class <br> name | Outdoor temp. <br> Range | Days | Corresponding <br> indoor readings |
| :--- | :--- | :--- | :--- | :--- |
| 1 | Coldest | $[-7.4$ to 7.00$]$ | 104 | Subset $1(104$ seq. $)$ |
| 2 | Cold | $[7.01$ to 15.1$]$ | 86 | Subset $2(86$ seq. $)$ |
| 3 | Hot | $[15.11$ to 22$]$ | 93 | Subset 3 (93 seq.) |
| 4 | Hottest | $[22.01$ to 38.8$]$ | 82 | Subset $4(82$ seq. $)$ |

This classification is done not only by matching with the seasons, but also by balancing the amount of data for each subset. Table 2 shows that class 1 typically represents the winter readings and class 4 the summer. The other two classes (class 2 and class 3 ) represent mainly from the spring and autumn seasons. However, depending on the dynamics of the data, the class or outdoor temperature range can be set to different numbers.

## 5. Numerical application

IOHMM uses the transition probability to predict the future events. Therefore, the training sequences should have information on transitions between one state to another state. Otherwise, the transition matrix cannot be learned correctly and will not work effectively. Therefore, all the sequences that do not have any/enough transitions are removed from the training set.

### 5.1 Training result

IOHMM training delivers four transition matrices according to four inputs. The inputs are already mentioned earlier which represent four different classes. For example:
Class 1 is represented by the first estimated transition matrix:

$$
\left(\begin{array}{llll}
0.000 & 0.467 & 0.164 & 0.369 \\
0.031 & 0.969 & 0.000 & 0.000 \\
0.001 & 0.001 & 0.285 & 0.713 \\
0.000 & 0.025 & 0.476 & 0.499
\end{array}\right) \text {, and }
$$

The correspondant emission matrix:

$$
\left(\begin{array}{lllll}
0.505 & 0.000 & 0.494 & 0.000 & 0.000 \\
0.693 & 0.000 & 0.245 & 0.058 & 0.004 \\
0.001 & 0.999 & 0.000 & 0.000 & 0.000 \\
0.001 & 0.999 & 0.000 & 0.000 & 0.000
\end{array}\right)
$$

These two matrices represent the complete parameters of IOHMM given the input $U=1$. The training method gives three more sets of (transition, emission) parameters for $U=2,3$, and 4 .

### 5.2 Testing result: on a single test sequence

The selected test sequence is converted to an incomplete sequence. Only the first half of the sequence is given to the model to predict the second half. The result is then compared to the original sequence (see Figure 7).


Figure 7. Comparison between predicted and original emitted (temperature) sequence

The Viterbi algorithm is used to predict the maximum path given the half (test) sequence, after which the model uses the transition and the emission matrices to predict the rest of the sequence elements following the Markov property.

## The details result:

Test sequence length: 47; Number of predicted elements: 49; Wrong prediction: 8 ; Performance: about $83.67 \%$ accuracy

### 5.3 Cross validation: Leave one out (LOO)

One sequence is just one instance of the reality which does not justify much of the model performance. Therefore, a cross validation can be useful and more trustworthy approach to validate the model performance on the dataset.

The Cross-validation is an experiment to analyze whether the predictive performance of a model declines significantly or does not when applies to new relevant data (Berrar, 2019). There are several popular cross-
validation methods used in literature articles. In this paper, the leave-one-out (LOO) method is used to show the belief over the model performance.

LOO is such a validation technique where a random data sample is set aside for testing, and the rest is used to train the model (Figure 8). The data samples (sequence for each day) are assumed as independent entities.


Case \#2

## Case \#n

Figure 8. Leave-one-out cross validation [4]
This method is illustrated on a dataset with $n=30$ cases. Each case is used in turn as a single retained test case. The model was built using the remaining $n-1$ cases.
The details LOO result:
It follows the similar technique as 4.1 , only with 30 times:

Number of sequences $=30$; Number of trainings $=$ 30; Number of tests $=30$; Prediction performance $=$ $\mathbf{7 8 . 4 \%}$; The error rate $=\mathbf{2 1 . 6 \%}$

### 5.4 Performance Comparison

Two performance tables (Tables 3 and 4) are presented, which are calculated by applying the LOO cross-validation method.

Table 3 represents the impact of different number of parameters of IOHMM on the performance with the same training and testing sequences. Choosing the right number of hidden states and discrete symbols is always a challenge for IOHMM-based model (Shahin, 2020).

We conducted an experiment where the number of hidden states was set as 3 to 5 and discrete symbols as 3 to 6 . This experiment could be extended to more parameter variations but based on the amount of data and model performance, we decided to go with such numbers.

There is a relationship between the number of hidden states and the number of transitions between discrete symbols that can be observed. Sometimes, fewer hidden states can move quickly from one temperature level to another by skipping several transitions, or sometimes more hidden states can be cause of misleading predictions because there are not enough transitions dedicated to each hidden state in the dataset. For example, version numbers 1 and 5 (Table 3) show the lowest performance because the observations
have only three discrete symbols. When the number of hidden states increases to 4 , the error rate increases because the additional hidden states do not have enough transitions in the discrete form of the data (with only three symbols).

On the other hand, when the number of discrete symbols increases to 5 or more (versions 7 and 8 ) with 4 hidden states, the model performs well because in this case there are sufficient transitions between symbols for all 4 hidden states.

However, IOHMM sometimes performs well with a high number of parameters, but this also increases the complexity of building and processing the model. For example, adding one discrete symbol to version 8 (compared to the version 7) is equivalent to adding 4 symbols (for four model inputs). Therefore, it needs to be carefully decided whether it is worth adding four additional parameters to improve the accuracy of (79.03 - 78.4) 0.63\%.

Another example can be given here in relation to more parameters that can badly influence the model performance is version 9 . Even though, it has more states and discrete variables but does not perform well. The reason behind this is the lack of data amount and poor model training. There are not enough transition and emission events exists in the data that support the extra parameters for version 9 . So, it could be tricky to raise the number of parameters.

Therefore, this experiment (Table 3) is performed on the same dataset in order to find an appropriate model version. In this case, IOHMM version 7 (with 4 hidden states and 5 discrete symbols) seems to be the most suitable model version for the given dataset.

Table 3. Model Performance for Different Settings

| Version <br> No | Number <br> of <br> Hidden <br> States | Number <br> of <br> Discrete <br> Symbols | Accuracy <br> rate | Error <br> rate |
| :--- | :--- | :--- | :--- | :--- |
| 1 | 3 | 3 | $61.38 \%$ | $38.62 \%$ |
| 2 | 3 | 4 | $69.21 \%$ | $30.79 \%$ |
| 3 | 3 | 5 | $68.00 \%$ | $32.00 \%$ |
| 4 | 3 | 6 | $71.05 \%$ | $28.95 \%$ |
| 5 | 4 | 3 | $56.98 \%$ | $43.02 \%$ |
| 6 | 4 | 4 | $76.65 \%$ | $23.35 \%$ |
| $\mathbf{7}$ | $\mathbf{4}$ | $\mathbf{5}$ | $\mathbf{7 8 . 4 \%}$ | $\mathbf{2 1 . 6 \%}$ |
| $\mathbf{8}$ | $\mathbf{4}$ | $\mathbf{6}$ | $\mathbf{7 9 . 0 3 \%}$ | $\mathbf{2 0 . 9 7 \%}$ |
| 9 | 5 | 6 | $74.01 \%$ | $\mathbf{2 5 . 9 9 \%}$ |

Once the model structure is fixed, the next experiment compares the performance of proposed version of IOHMM with an HMM-based predictor.

Table 4 shows the performance of the proposed IOHMM compared to the HMM as the baseline prediction point. The HMM uses the normal ML algorithm that uses all the data for model training, and it predicts the average temperature of all the dynamics on the dataset. On the other hand, IOHMM uses an adapted ML algorithm that considers the inputs in model development. Thus, it predicts the exact temperature for a given input signal that represents different classes mentioned in Table 2.

Table 4. Average temperature prediction vs IOHMM prediction

| Model | Number <br> of <br> Sequence | Number <br> of <br> Training <br> and <br> testing | Accuracy <br> rate | Error <br> rate |
| :--- | :--- | :--- | :--- | :--- |
| HMM | 30 | 30 | $69.07 \%$ | 30.93 <br> $\%$ |
| IOHMM | 30 | 30 | $\mathbf{7 8 . 4 \%}$ | $\mathbf{2 1 . 6 \%}$ |

IOHMM first obtains inputs to index which class the test sequence comes from. It then applies the corresponding model parameters to predict the temperature. It provides more efficient and plausible results compared to a common ML-based predictor (see Table 4).

The model performance can be improved by more precise data along with the data uncertainty handling approach.

### 5.5 Example: The preference algorithm

Suppose an employee comes to the office at 9:00 a.m. He wants to find a cooler space for a few hours e.g., 5 hours, and then he leaves the office for the day. The proposed method starts with obtaining preferences from the employee and then gives the solution in four steps.
Step 01: Predict the current thermal conditions at all available workstations (including different rooms).

Step 02: Identify the workstations that match the preferred temperature (or relatively cooler space).

Step 03: Predict the temperature of each selected workstation for the next five hours and checks if they remain cooler until then. If not, then it suggests the closest available option to the employee.

Step 04: Finally, when a workstation is occupied, the method sends a signal to the database by indexing the workstation to be occupied.

Noted that, if there is no available workstation, the method notifies the employee immediately that all workstations are occupied, or it asks to change the preferences.

## 6. Conclusion

In this paper, a probabilistic input-output Hidden Markov Model is proposed to predict indoor thermal conditions which allow choosing the desirable seating spaces based on past and the most probable future events. In the context of the IOHMM, the proposed approach currently considers multiple input modes, representing different dynamics of outdoor temperature and the corresponding effects on indoor thermal conditions. The model not only predicts the next events but also simulates the most probable (upcoming) thermal conditions until a given time considering the inputs. Multiple combinations between different variables of the indoor environment are handled as inputs of the model for predicting the future events that are close to reality which later compared to the original reading in the performance validation. A numerical application is implemented to demonstrate the functionality of the model. A benchmark is presented where the model is executed with different number of parameters but on the same data set to determine the appropriate number of hidden states and discrete symbols for the proposed IOHMM. The performance of the model is demonstrated by leaving one out a crossvalidation technique.

By adding more inputs and corresponding parameters, the proposed IOHMM can be used to simulate multiple dynamics of thermal conditions arising from different causes in the same room or even an entire office. The model also allows to monitor an entire floors or small buildings with only a single version of it. The training and testing algorithms are suitable for considering multiple observable outputs in model development, enabling the study of multiple observations (barometric pressure, humidity, etc.) in the prediction methods. This not only ensures the efficiency of the results, but also provides the opportunity for future work with multiple input and output transformations.

## Acknowledgement

The authors would like to acknowledge funding by EUDP (Grant, n. 64018-0558) and Elforsk (Grant, n. 351-034)

## References

Alawadi, Sadi, et al. "A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings." Energy Systems (2020): 1-17.
Alrazgan, Abdullah, et al. "Learning occupancy prediction models with decision-guidance query language." 2011 44th Hawaii International Conference on System Sciences. IEEE, 2011.
Arendt, Krzysztof, et al. "Comparative analysis of white-, gray-and black-box models for thermal simulation of indoor environment: Teaching building case study." Proceedings of the 2018 Building Performance Modeling Conference and SimBuild co-organized by ASHRAE and IBPSA-USA, Chicago, IL, USA. 2018.
Berrar, Daniel. "Cross-Validation." (2019): 542-545.
Chafi, Maral Babapour, Mette Harder, and Christina Bodin Danielsson. "Workspace preferences and nonpreferences in Activity-based Flexible Offices: Two case studies." Applied ergonomics 83 (2020): 102971.
Ellis, Carl, Mike Hazas, and James Scott. "Matchstick: A room-to-room thermal model for predicting indoor temperature from wireless sensor data." Proceedings of the 12th international conference on Information processing in sensor networks. 2013.
Fang, Zhen, et al. "multi-zone indoor temperature prediction with LSTM-based sequence to sequence model." Energy and Buildings 245 (2021): 111053.
Hietaharju, Petri, Mika Ruusunen, and Kauko Leiviskä. "A dynamic model for indoor temperature prediction in buildings." Energies 11.6 (2018): 1477.
Hu J, Wang Y, Zhang Y. Iohmm for location prediction with missing data. In Data Science and Advanced Analytics (DSAA). IEEE International Conference on 2015 Oct 19 (pp. 1-10).
Huang, Yongchao, Hugh Miles, and Pengfei Zhang. "A Sequential Modelling Approach for Indoor Temperature Prediction and Heating Control in Smart Buildings." arXiv preprint arXiv:2009.09847 (2020).
Jens, Krister, and Jay Sterling Gregg. "How design shapes space choice behaviors in public urban and shared indoor spaces-A review." Sustainable Cities and Society (2020): 102592. Just A, Bernier O, Marcel S. HMM and IOHMM for the recognition of mono-and bi-manual 3D hand gestures. IDIAP; 2004.
Khan, Donya Sheikh, Jakub Kolarik, and Peter Weitzmann. "Design and application of occupant voting systems for collecting occupant feedback on indoor environmental quality of buildings-a review." Building and Environment (2020): 107192.
LeMay, Michael, Jason J. Haas, and Carl A. Gunter. "Collaborative recommender systems for building automation." 2009 42nd Hawaii International Conference on System Sciences. IEEE, 2009.
Lu, Siliang, et al. "Data-driven simulation of a thermal comfort-based temperature set-point control with ASHRAE RP884." Building and Environment 156 (2019): 137-146.

Ma, Nan, et al. "Adaptive behavior and different thermal experiences of real people: A Bayesian neural network
approach to thermal preference prediction and classification." Building and Environment 198 (2021): 107875.

Ourston, Dirk, et al. "Applications of hidden markov models to detecting multi-stage network attacks." 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the. IEEE, 2003.

Pathak, Jyotishman, et al. "Condition data aggregation with application to failure rate calculation of power transformers." Proceedings of the 39th annual Hawaii international conference on system sciences (HICSS'06). Vol. 10. IEEE, 2006.
Peng, Yuzhen, et al. "Case study review: Prediction techniques in intelligent HVAC control systems." 9th International Conference on Indoor Air Quality Ventilation and Energy Conservation in Buildings (IAQVEC 2016). 2016.

Perić, Nikica, et al. "Prediction of User-defined Room Temperature Setpoints by LSTM Neural Networks." 2021 23rd International Conference on Process Control (PC). IEEE, 2021.
Rabiner, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." Proceedings of the IEEE 77.2 (1989): 257-286.
Rahaman, Mohammad Saiedur, et al. "OccuSpace: Towards a robust occupancy prediction system for activity-based workplace." 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE, 2019.
Rahaman, Mohammad Saiedur, et al. "Seating preference analysis for hybrid workplaces." arXiv preprint arXiv :2007.15807 (2020).
Sanandaji, Borhan M., He Hao, and Kameshwar Poolla. "Fast regulation service provision via aggregation of thermostatically controlled loads." 2014 47th Hawaii International Conference on System Sciences. IEEE, 2014.

Shahin, Kamrul Islam, Christophe Simon, and Philippe Weber. "Estimating IOHMM parameters to compute remaining useful life of system." Proceedings of the 29th European Safety and Reliability Conference, Hannover, Germany. 2019.
Shahin, Kamrul Islam. Dynamic probabilistic graphical model applied to the system health diagnosis, prognosis, and the remains useful life estimation. Diss. Université de Lorraine, 2020.
Sood, Tapeesh, Patrick Janssen, and Clayton Miller. "Spacematch: Using environmental preferences to match occupants to suitable activity-based workspaces." arXiv preprint arXiv:2006.09570 (2020).
Tariq, Zaid Bin, et al. "Experimental evaluation of data-driven predictive indoor thermal management." Proceedings of the Tenth ACM International Conference on Future Energy Systems. 2019.
Weber P, Simon C. Benefits of Bayesian Network Models. John Wiley \& Sons; 2016 Aug 23
Bengio Y, Frasconi P. An input output HMM architecture. In Advances in neural information processing systems 1995 (pp. 427-434)

