

建筑用户在室行为预测新方法

俞准[†], 周亚苹, 李郡, 黄余建, 张国强

(湖南大学 土木工程学院, 建筑安全与环境国际联合研究中心, 湖南 长沙 410082)

摘要:准确预测建筑用户在室行为可显著提高建筑能耗模拟精度,并进一步帮助建筑设计及运行控制优化.当前进行在室行为预测时所采用的主要是基于隐马尔可夫链方法的数学模型,该模型考虑了在室行为的时间关联性,可平稳有效地预测在室行为.然而现有隐马尔可夫模型难以准确描述在室行为动态变化规律以及在室行为与可观测参数之间的关联,降低了模型预测精度.针对该问题,本文提出一种基于状态转移的时变隐马尔可夫模型.该模型采用时变状态转移概率矩阵量化不同时刻在室行为的动态变化特征及关联,同时该模型基于状态转移计算可观测参数的概率分布以定量描述在室行为对可观测参数的影响.本文采用比利时某办公室在室行为数据库进行了相关建模和验证,结果表明该模型可更有效地捕捉在室状态变化,从而提高了在室行为预测精度.

关键词:在室行为;人行为;预测;隐马尔可夫模型;建筑模拟

中图分类号:TU201.5

文献标志码:A

A New Approach for Building Occupancy Prediction

YU Zhun[†], ZHOU Yaping, LI Jun, HUANG Yujian, ZHANG Guoqiang

(National Center for International Research Collaboration in Building Safety and Environment,
College of Civil Engineering, Hunan University, Changsha 410082, China)

Abstract: Accurate prediction of occupancy in buildings can significantly improve the performance of building energy simulation and further facilitate building design and system operation. Considering the temporal dependency of occupancy, Hidden Markov Model has been widely used to effectively predict occupancy behavior. However, the traditional Hidden Markov model that uses time-independent transition probability matrix is difficult to accurately describe the dynamic variation of occupancy as well as its correlation with environmental parameters. Such a model would greatly reduce occupancy prediction accuracy. To address this issue, an inhomogeneous Hidden Markov Model based on state transition was proposed. In this model, time-dependent transition probability matrices were calculated to capture the temporal dependency of occupancy at different time periods. Meanwhile, probability distribution of environmental parameters was calculated based on state transition instead of state only, aiming at rationally describing the correlation between occupancy and environmental parameters. The method was applied to predict the occupancy of a Belgian office. The results demonstrated that the effectiveness of the proposed approach and the prediction accuracy were improved significantly.

Key words: occupancy; occupant behavior; prediction; hidden Markov model; building simulation

* 收稿日期:2018-10-18

基金项目:国家自然科学基金资助项目(51408205), National Natural Science Foundation of China(51408205);中央高校基本科研业务费专项资金资助项目(2014-021), The Fundamental Research Funds for the Central Universities(2014-021);湖南省建筑节能与环境控制关键技术协同创新中心项目(201736370287), Hunan Collaborative Innovation Centre of Building Energy Conservation & Environmental Control(201736370287)

作者简介:俞准(1980—),男,安徽芜湖人,湖南大学副教授,博士

[†] 通讯联系人, E-mail: zhunyu@hnu.edu.cn

建筑用户在室行为主要包括在室状态和时长、用户位置、在室人数和用户活动^[1]. 对其进行准确预测并进一步与建筑能耗模拟软件(如 EnergyPlus 和 DeST)集成,可显著提高建筑能耗模拟精度,同时有助于建筑性能评估和建筑系统运行优化控制^[2-5]. 现有在室行为预测方法主要是在测量多种室内环境相关数据(如 CO₂ 浓度和温湿度)基础上,通过选取合适的环境特征参数建立相应模型,所采用的数学方法主要是隐马尔科夫法. 例如, Dong 等人基于温湿度、CO₂ 浓度和噪音数据,采用该方法建立了隐马尔科夫模型(Hidden Markov Model, HMM)对办公室的在室人数进行预测^[6]. Candanedo 等人基于 CO₂ 浓度、温湿度和照度数据,采用 HMM 对住户的在室状态进行预测^[7]. 上述研究表明 HMM 通过引入状态转移概率矩阵参数,定量描述了在室行为的时间关联性,因此可平稳有效地预测在室行为. 然而现有 HMM 模型还存在一定局限性,主要表现在以下两点:第一,忽略了不同时刻在室行为动态变化的差异性,导致在室行为动态变化规律与实际存在较大差异;第二,无法考虑过去的在室行为对当前可观测的环境特征参数的影响,导致在基于环境数据预测在室行为时难以得到满意的精确度.

为了解决上述问题,本文在 HMM 的基础上,建立了一种基于状态转移的时变隐马尔科夫模型(Inhomogeneous Hidden Markov Model based on State Transitions, TIHMM)对在室行为进行预测,并且采用比利时某办公室在室行为数据库对该预测模型进行了验证.

1 在室行为预测方法

1.1 TIHMM 模型基本原理

HMM 是一种包含隐含状态和可观测参数的概率模型,其目的是通过可观测参数推断隐含状

态^[8-10]. 模型假设隐含状态(如在室状态或人数)的变化过程具有马尔科夫特性,即任一时刻的隐含状态仅与前一状态及转移概率有关,可观测参数(如环境特征参数)的变化规律则由基于各隐含状态的概率分布进行描述. HMM 可由下列 5 个参数描述:

1) $S = \{s^1, s^2, \dots, s^M\}$ 为隐含的在室行为状态集,其中 M 为在室行为的状态数目.

2) $X = \{x_1, x_2, \dots, x_n\}$ 为可观测的环境特征参数集,其中 N 为特征参数数目.

3) $\pi = \{\pi^i\}$ 为各隐含状态的初始概率分布,其中 $\pi^i = P(s^i), 1 \leq i \leq M$.

4) $A = \{a^{ij}\}$ 为隐含状态转移概率矩阵,其中, $a^{ij} = P(s_{t+1}^j | s_t^i), 1 \leq i, j \leq M$.

5) $B = \{b^j\}$ 为可观测参数的输出概率分布,其中, $b^j = P(x_1, x_2, \dots, x_n | s_t^j), 1 \leq j \leq M$.

为克服现有 HMM 模型预测在室行为时的缺陷(如前文所述),本文提出对其进行如下改进并建立 TIHMM 模型:1)采用时变状态转移概率矩阵,代替 HMM 中不随时间变化的状态转移矩阵,以量化不同时刻在室行为的动态变化特征及关联,建立不均匀隐马尔科夫模型(Inhomogeneous Hidden Markov Model, IHMM);2)基于状态转移计算环境特征参数的输出概率分布,代替 HMM 中基于状态计算的输出概率分布,以定量分析在室行为变化对当前环境相关观测量的影响. TIHMM 改进的参数具体描述如下:

1) $A_t = \{a_{i+1,t}^j\}$ 为状态转移概率矩阵,其中 $a_{i+1,t}^j = P(s_{t+1}^j | s_t^i), 1 \leq i, j \leq M$, 表示 t 时刻状态为 $s^i, t+1$ 时刻状态为 s^j 的概率.

2) $B = \{b_t^j\}$ 输出概率分布,其中 $b_t^j = P(x_1, x_2, \dots, x_n | s_t^j, s_{t-1}^i), 1 \leq i, j \leq M$, 表示 $t-1$ 时刻状态为 s^i, t 时刻状态为 s^j 的条件下, t 时刻环境特征参数值的概率.

TIHMM 模型建立基本原理如图 1 所示.

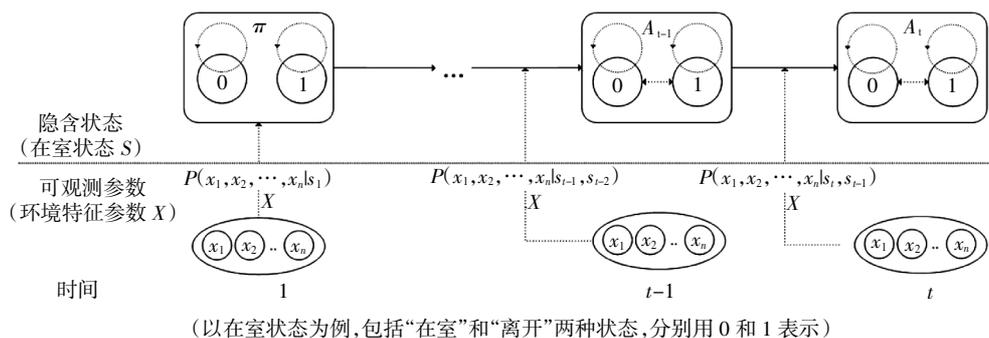


图 1 TIHMM 示意图

Fig.1 Schematic illustration of the TIHMM

1.2 参数学习

参数学习是建立 TIHMM 模型的关键,其主要目的是利用在室行为相关数据确定模型的基本参数,包括初始概率、状态转移概率和输出概率。

1)初始概率 $P(s_i^j)$ 和状态转移概率 $P(s_{t+1}^j | s_t^i)$

由于本文所提出的模型是基于不均匀隐马尔可夫模型所建立的,合理确定状态转移的时间步长是保证模型预测性能的前提.已有研究表明采用 10 min 模拟步长可有效捕捉到在室行为的变化,且采用与模拟步长一致的状态转移时间步长具有最优预测效果^[11-12].因此本文选取状态转移时间步长为 10 min,并基于在室行为的实际数据学习模型的初始概率和状态转移概率,计算方法如下:

$$P(s_i^j) = \frac{n_i^j}{\sum_{i=1}^M n_i^j} \quad (1)$$

$$P(s_{t+1}^j | s_t^i) = \frac{n_{t+1,t}^{ji}}{\sum_{j=1}^M n_{t+1,t}^{ji}} \quad (2)$$

式中: $n_{t+1,t}^{ji}$ 表示 $t, t+1$ 时刻状态分别为 s_t^i, s_{t+1}^j 的数目; n_i^j 表示 t 时刻状态为 s_i^j 的数目; M 表示在室行为的状态数目; $t = 1, 2, \dots, 144$.

2)输出概率 b_i^j

本文采用混合高斯模型计算不同状态转移的输出概率分布,计算公式如式(3)所示^[6].以在室状态为例,状态转移分为以下 4 种:持续为“离开”状态、持续为“在室”状态、从“离开”变化为“在室”状态以及从“在室”变化为“离开”状态。

$$b_{ji}^j(x) = P(x | s_{t+1}^j, s_t^i) = \sum_{k=1}^K c_{jim} N(x; \mu^{jim}, \Sigma^{jim}) \quad (3)$$

式中: x 表示环境特征参数集; s_{t+1}^j 表示 $t+1$ 时刻的在室行为状态; s_t^i 表示 t 时刻的在室行为状态; K 表示高斯模型个数; c_{jim} 表示各个高斯模型所占权重; $N(x; \mu^{jim}, \Sigma^{jim})$ 为单个高斯模型的概率分布。

1.3 在室行为预测

在确定 TIHMM 的模型基本参数之后,为实时预测在室行为,本文基于初始概率、状态转移概率以及输出概率,在可观测的环境特征参数序列 $x_{1:t}$ 下,通过前向算法递归计算得到最大后验概率,从而推断出最可能出现的在室行为状态 $s_t^{[13]}$:

$$\operatorname{argmax}_s P(s_t | x_{1:t}) \quad (4)$$

具体预测算法如表 1 所示。

表 1 在室行为预测算法

Tab.1 Occupancy prediction algorithm

采用最大后验概率联合前向算法对在室行为进行一天实时预测

输入: $\pi = \{\pi_i\}, A_i = \{a_{i,j}^i\}, B = \{b_i^j\}, i, j = 1, 2, \dots, m$

$t = 1, 2, \dots, L, L = 144$

输出:在室行为 Q

初始化: $t = 1$

1: $a_i^i = \pi_i b_i^i$

2: $Q_t = \operatorname{argmax}_i a_i^i$

循环:

3: for $t = 1$ to $L-1$ do

4: $a_{t+1}^j = \sum_{i=1}^m (b_{t+1,t}^{ji} a_t^i)$

5: $Q_{t+1} = \operatorname{argmax}_j a_{t+1}^j$

6: end for

7: return Q

2 特征选择

考虑到环境数据中存在异常值,直接使用原始环境数据学习模型参数将会导致模型预测精度显著降低.同时,原始环境数据中包含的在室行为相关信息相对较少,需从原始信息中提取有用特征.因此,本文在建模之前首先对数据进行平均化处理以降噪,在此基础上建立由在室行为相关环境特征参数组成的特征参数集,并通过相对信息增益度量不同环境特征参数与在室行为之间的相关程度,从而对特征参数进行排序以进一步筛选确定模型最佳输入特征参数组合,具体步骤如下:

1)平均化处理.对于数据库中的原始环境数据,根据式(5)计算平均值。

$$X_k = \frac{R_k + \dots + R_{k-l+1}}{l} \quad (5)$$

式中: X_k 表示 k 时刻平均化处理后的环境参数; R_k 表示原始环境参数; l 表示模拟步长, $l = 10 \text{ min}$ 。

2)特征参数集建立.本文建立包括原始特征、一阶差分、二阶差分和一阶移动差分的特征参数集,如表 2 所示。

表 2 特征参数集

Tab.2 List of features

特征名称	描述
原始特征	$X(k)$
一阶差分	$fd(k) = X(k) - X(k-1)$
二阶差分	$sd(k) = fd(k) - fd(k-1)$
一阶移动差分	$fds(k) = X(k) - X(k-2)$

3)相对信息增益法. 相对信息增益是一种基于信息熵和条件熵评估特征参数(如环境特征参数)给分类(如在室行为)带来信息量的方法^[14]. 其中信息熵是信息空间的不确定性或随机性程度的度量,熵越小,表示信息空间概率分布越均匀. 条件熵表示观测某个变量之后信息空间的不确定度. 计算步骤如下.

①计算在室行为 y 的信息熵 $H(y)$:

$$H(y) = -\sum_{i=1}^m P(y_i) \log_2 P(y_i) \quad (6)$$

②计算环境特征参数 x 对在室行为 y 的条件熵 $H(y|x)$:

$$H(y|x) = -\sum_{j=1}^n P(x=x_j) H(y|x=x_j) \quad (7)$$

③计算 y 与 x 之间的信息增益 IG:

$$IG(y,x) = H(y) - H(y|x) \quad (8)$$

④计算相对信息增益 RIG:

$$RIG(y,x) = \frac{IG(y,x)}{H(y)} \times 100\% \quad (9)$$

式中: M 表示在室行为包含的状态数目; $P(y_i)$ 表示数据库中在室行为属于第*i*类状态的概率; D 表示将连续的环境特征参数离散化后的子集数目; $P(x=x_j)$ 表示数据库中环境特征参数 x 属于第*j*类子集的概率.

3 案例研究

3.1 数据库

为验证所提出方法的可行性,本文采用比利时某双人办公室在室行为数据库建立在室状态预测模型并对预测结果进行评估^[15].该数据库连续9d采集了某办公室的温度、湿度、照度、CO₂浓度和用户在室状态(包括“在室”和“离开”两种状态,分别用1和0表示),采集间隔时间为1min.

3.2 评估指标

在室状态预测可视作一种分类预测问题,即根据环境特征参数推断用户“在室”和“离开”两种状态是否被正确分类.因此,本文通过分类预测准确度评价模型整体预测性能,计算公式如式(10)所示.与此同时,由于模型预测结果可能出现两类错误(“在室”误判为“离开”状态、“离开”误判为“在室”状态),若将模型应用于优化控制建筑设备系统时(如照明系统和暖通空调系统),这两类错误将分别导致室内舒适度降低和不必要的能源浪费.因此,

本文对“在室”和“离开”两种状态的预测精度分别进行检验,计算公式如式(11)(12)所示.

模型整体预测精度:

$$T_m = (A+D)/(A+B+C+D) \quad (10)$$

“在室”预测精度:

$$T_p = (A)/(A+B) \quad (11)$$

“离开”预测精度:

$$T_n = (D)/(C+D) \quad (12)$$

式中: A 表示实际和预测状态均为“在室”的数目; B 表示误判“离开”为“在室”状态的数目; C 表示误判“在室”为“离开”状态的数目; D 表示实际和预测状态均为“离开”的数目.

3.3 结果和讨论

3.3.1 特征参数选择分析

表3给出了不同环境特征参数与在室状态之间的相对信息增益计算结果.从表中可知,CO₂浓度、照度和温度的相对信息增益较大,即这3类环境因素与在室状态的相关性较大.同时,相对于其他特征类型,二阶差分与在室状态相关性最小.本文最终选择的模型输入特征参数包括CO₂浓度的原始特征、一阶差分和一阶移动差分以及照度和温度的原始特征.

表3 特征参数的相对信息增益

Tab.3 Relative information gain values of the features

环境因素	相对信息增益			
	原始	一阶差分	一阶移动差分	二阶差分
CO ₂ 浓度	0.665 4	0.556 5	0.572 3	0.326 4
照度	0.889 4	0.423 1	0.456 5	0.464 5
温度	0.509 8	0.108 5	0.123 9	0.080 0
湿度	0.180 7	0.197 3	0.221 8	0.129 9
热湿比	0.340 0	0.344 0	0.348 5	0.162 7

3.3.2 模型比较

为验证提出方法的有效性,本文基于上述环境特征参数建立了TIHMM在室行为预测模型,并与传统HMM和IHMM分别进行了比较.

图2~图4分别给出了采用TIHMM、IHMM和HMM预测的在室状态变化曲线图.由图可见,相对于IHMM和HMM,TIHMM的预测结果更加符合实际的在室状态变化曲线,尤其是用户最先到达时间和最后离开时间这两项信息.考虑到在进行建筑系

统运行控制优化时,系统启停时间主要取决于用户最先到达和最后离开时间,对这两项信息进行准确预测有助于提高系统运行效率,具有显著的实际意义^[6].从图中还可看到,当发生的状态变化持续一段时间时(如图中 13:00~14:00),采用 TIHMM 预测的状态变化与实际变化相同,而 IHMM 和 HMM 的预测结果均出现了不同程度的延迟.这证明和 IHMM 以及 HMM 相比,TIHMM 由于采用了不均匀状态转移概率矩阵,且同时考虑了在室状态变化对当前环境特征参数的影响,因此减弱了在室行为的随机特性和环境数据的延迟特性(如用户呼出的 CO₂ 均匀扩散到室内需要一定响应时间)对在室状态预测的影响程度,从而能够更准确地反映在室状态的动态变化规律.

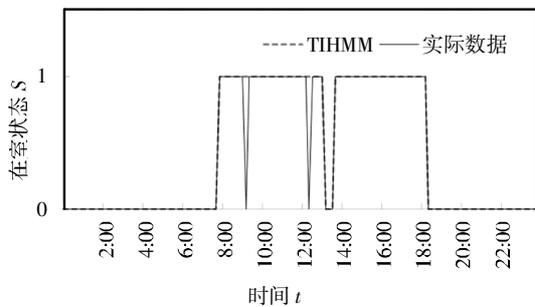


图 2 TIHMM 预测在室状态结果图

Fig.2 Occupancy prediction results based on TIHMM

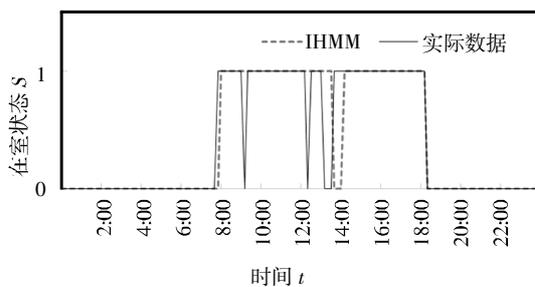


图 3 IHMM 预测在室状态结果图

Fig.3 Occupancy prediction results based on IHMM

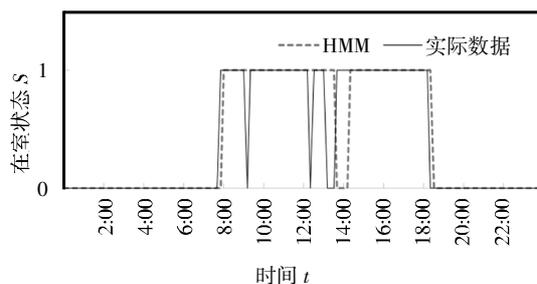


图 4 HMM 预测在室状态结果图

Fig.4 Occupancy prediction results based on HMM

值得注意的是,对 8:00~10:00 和 12:00~13:00 两个期间出现的短时间在室状态变化,TIHMM、IHMM 和 HMM 这 3 种模型均未能进行有效预测.一个可能的原因是该类变化产生的环境特征参数变化小,与不发生状态变化的环境特征参数相近.需要指出的是,在实际应用过程中,为避免系统的频繁启停,该类预测误差变化往往被忽略^[7].

为进一步比较模型预测性能,表 4 给出了这 3 种模型的整体、“在室”状态和“离开”状态的预测精度.结果表明,3 种检验指标下 TIHMM 的预测效果均为最优.同时,3 种模型对“离开”状态的预测精度均高于对“在室”状态的预测精度,其主要原因是在非工作时间段(如图 2~图 4 中 12:00~7:00),用户在室状态相对稳定且均为“离开”,因此将实际“在室”状态误判为“离开”状态的可能性较小;在工作时间段,由于模型难以捕捉到短时间的状态变化,且该类变化通常是“离开”状态,因此,将实际“离开”状态误判为“在室”状态的可能性较大.

表 4 不同模型预测精度对比

Tab.4 Comparison between the prediction accuracy of the three models

评估指标	预测精度		
	TIHMM	IHMM	HMM
整体预测精度	0.981 0	0.954 3	0.943 0
“在室”预测精度	0.947 9	0.915 8	0.905 2
“离开”预测精度	1.0	0.976 2	0.964 3

4 结论和未来展望

本文提出了一种基于 TIHMM 预测在室行为的新方法.该方法从两方面克服了现有 HMM 的缺陷:第一,通过采用随模拟步长变化的状态转移概率矩阵,可更准确地反映在室行为动态变化规律;第二,基于状态转移计算环境特征参数的输出概率,从而量化了过去在室行为对当前环境特征参数的影响.本文采用包括环境和在室状态数据的比利时某办公室在室行为数据库,分别建立了 TIHMM、HMM 和 IHMM 对在室状态进行预测,并对这 3 种模型从趋势变化和总体性能两方面进行了比较.结果显示相对于其它两种模型,TIHMM 预测结果更符合实际在

室状态变化趋势,预测精度更高。

本文研究主要针对多人单区域进行在室状态预测,在此基础上,未来应进一步建立多人多区域的在室行为预测模型,以获取在室人数和用户位置等更全面的在室行为相关信息,并将其有效应用于建筑系统设计及运行控制优化中。

参考文献

- [1] YAN D, O'BRIEN W, HONG T, *et al.* Occupant behavior modeling for building performance simulation: Current state and future challenges[J]. *Energy & Buildings*, 2015, 107:264—278.
- [2] 周亚苹,俞准,李郡,等. 建筑用户在室行为测量方法及预测模型综述[J]. *暖通空调*, 2017, 47(9):11—18.
ZHOU Y P, YU Z, LI J, *et al.* Review of measuring methods and prediction models of building occupancy [J]. *Journal of HV&AC*, 2017, 47(9):11—18. (In Chinese)
- [3] YU Z, HAGHIGHAT F, FUNG B C M. Advances and challenges in building engineering and data mining applications for energy-efficient communities[J]. *Sustainable Cities & Society*, 2016, 25: 33—38.
- [4] GUO X, TILLER D K, HENZE G P, *et al.* The performance of occupancy-based lighting control systems: A review[J]. *Lighting Research & Technology*, 2010, 42(1):415—431.
- [5] 杨昌智,文洁,蒋新波. 热湿环境参数对 PMV 及空调能耗的影响研究 [J]. *湖南大学学报 (自然科学版)*, 2015, 42(1):104—108.
YANG C Z, WEN J, JIANG X B. Study on the influence of thermal environmental parameters on PMV and energy consumption [J]. *Journal of Hunan University (Natural Sciences)*, 2015, 42(1): 104—108. (In Chinese)
- [6] DONG B, ANDREWS B, LAM K P, *et al.* An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network [J]. *Energy & Buildings*, 2010, 42(7):1038—1046.
- [7] CANDANEDO L M, FELDHEIM V, DERAMAIX D. A methodology based on Hidden Markov Models for occupancy detection and a case study in a low energy residential building [J]. *Energy & Buildings*, 2017, 148: 327—341.
- [8] GRIMMETT G R, STIRZAKER D R. *Probability and random processes second edition* [J]. Oxford Science Publications, 1992(1): 164—165.
- [9] 何磊,马伯祥,宗长富,等. 基于意图辨识的线控汽车紧急转向控制方法 [J]. *湖南大学学报 (自然科学版)*, 2014, 41(1):81—86.
HE L, MA B X, ZONG C F, *et al.* Emergency steering control based on driver steering intention recognition for steer-by-wire vehicle [J]. *Journal of Hunan University (Natural Sciences)*, 2014, 41(1): 81—86. (In Chinese)
- [10] RABINER L R. A tutorial on hidden markov models and selected applications in speech recognition[J]. *Readings in Speech Recognition*, 1990, 77(2):267—296.
- [11] ZHOU Y, YU Z, LI J, *et al.* The effect of temporal resolution on the accuracy of predicting building occupant behavior based on markov chain models[J]. *Procedia Engineering*, 2017, 205: 1698—1704.
- [12] LAM K P, HOYNCK M, DONG B, *et al.* Occupancy detection through an extensive environmental sensor network in an open-plan office building [C]//Eleventh International Ibpsa Conference. Glasgow, Scotland: IBPSA Conference, 2009: 1452—1459.
- [13] CHEN Z, ZHU Q, MASOOD M, *et al.* Environmental sensors based occupancy estimation in buildings via IHMM-MLR [J]. *IEEE Transactions on Industrial Informatics*, 2017, 99: 1—1.
- [14] MOZER M C. *The neural network house: an environment that adapts to its inhabitants*[R]. Boulder: University of Colorado Boulder, 1998: 110—114.
- [15] CANDANEDO L M, FELDHEIM V. Accurate occupancy detection of an office room from light, temperature, humidity and CO₂ measurements using statistical learning models [J]. *Energy & Buildings*, 2016, 112: 28—39.
- [16] CHEN Z, MASOOD M K, SOH Y C. A fusion framework for occupancy estimation in office buildings based on environmental sensor data[J]. *Energy & Buildings*, 2016, 133: 790—798.
- [17] GHOFRANI A, JAFARI M A. Distributed air conditioning control in commercial buildings based on a physical-statistical approach [J]. *Energy & Buildings*, 2017, 145: 106—118.