2016 IEEE International Conference on Automation Science and Engineering (CASE)
Fort Worth, TX, USA, August 21-24, 2016

Inferring Building Occupancy Based on Statistical Modeling of Multi-sensor Data

Bolun Zhang, Yuan Shen, and Lin Zhang

Abstract— **The building, being one of the largest energy consumers, accounts for over 40% of the global energy consumption. The traditional Heating, Ventilation and Air Conditioning (HVAC) and lighting systems cause considerable energy wastes in the building since they cannot adapt to the time-varying occupancy levels, which is expensive to measure with dedicated sensing systems. We propose an indirect method to estimate occupancy levels based on statistical modeling of environmental and energy data that can be measured by existing building sensor systems. The model can parameterize the empirical distribution of the environment parameters with small training data sets. Furthermore, statistical features are selected by the physical model to reduce the time cumulative effect of environment parameters. The proposed method is evaluated on a real testbed with CO_2 concentration and the total plugs power in the room. Results show that the proposed method can reduce the root mean squared error of the occupancy estimation by 33% compared to that using the empirical distribution.**

I. INTRODUCTION

Buildings contribute a major portion to the overall energy consumption of human being. With the rapid economy growth and contiguous environmental constraints, there are stronger demands to improve buildings energy efficiency. In the United States, building energy consumption accounts for 41% of the total energy consumption, in which nearly 62% is consumed by Heating, Ventilation and Air Conditioning (HVAC) system and 9% by lighting system [1]. Similarly, in the European Union countries, buildings are responsible for 40% of energy consumption [2]. Large amount of energy wastes is caused by the traditional control strategy of HVAC and lighting systems that does not adapt to the occupancy level variation.

In recent years, many schemes on improving control strategies for energy saving have proposed, and the key to these proposals is to estimate the occupancy levels in realtime [3][4]. The existing estimation methods can be generally categorized into two categories: direct and indirect methods.

The direct detection method uses dedicated devices such as RFID [5], camera [6] or passive infrared (PIR) sensors [7]. Besides the costly deployment that prevent these technologies being widely used, each has its own limitations. Camera-based occupancy detection is pretty intrusive and raises privacy concerns. In RFID-based detection solution, every occupant must always wear a RFID tag, which is impractical. Privacy-preserving and nonintrusive though, PIR sensors cannot provide precise occupancy levels since they can only provide binary information about the occupancy in a region. Moreover, when indoor people are frequently not moving for a long time when reading or working, PIR sensors will lost tracking occupants.

The indirect detection method infers occupancy levels based on relevant environmental parameters which are directly impacted by human activities [8]. The indirect detection can be further categorized into two types: the physical method [9][10] and the statistical inference method [11][12]. The physical method estimates occupancy levels by exploiting knowledge of the underlying physical laws, such as the mass-balance equation, dispersion process. However, it is a challenging task to establish reliable physical models in practice due to the environment diversity and the random behavior of humans. In the statistical inference method, the statistical correlations are established to connect environmental parameters to occupancy levels. This typically requires a lot of training data. Therefore, the statistical inference method may have a poor performance when training data are insufficient.

This paper proposes a scheme that combines the physical model and the statistical model. When large amount of training data is not available, the scheme uses statistical distribution instead of empirical distribution to deal with the training data. By analyzing the empirical distribution of environmental parameters, we choose suitable family of probability distribution functions (pdf) and calculate the corresponding parameters, which is capable of improving accuracy with limited training data. Our method is based on the two statistical inference methods: maximum a posteriori probability (MAP) estimation of Hidden Markov Model (HMM) and Multiple-Hypothesis Sequential Probability Ratio Test (MSPRT).

For privacy reasons, the proposed scheme will only detect the number of occupants, rather than differentiating individuals, as the number of occupants is sufficient for the energy-saving control for HVAC and lighting. Since CO_2 concentration is less affected by the environment and can reflect the change caused by human activity, CO_2 concentration is chosen as observed quantity. The total plugs power in the room is another observed quantity as it depends on human behavior. In view of time accumulation effect of CO_2, we extract features by physical model for inference. We evaluate our scheme on real data and compare its performance with Artificial Neutral Network (ANN), a state-of-the-art data driven method, the traditional HMM and MSPRT methods.

B. Zhang and Y. Shen are with Department of Electronic Engineering, Tsinghua University, Beijing, China, (email: zbl13@mails.tsinghua.edu.cn, shenyuan_ee@tsinghua.edu.cn).

L. Zhang is with Department of Electronic Engineering, Tsinghua University, Beijing, and Tsinghua-Berkeley Shenzhen Institute, Shenzhen, China(email: linzhang@tsinghua.edu.cn).

Results show that our scheme is more accurate than ANN, the traditional HMM and MSPRT. It is also demonstrated that the multi-sensor information fusion can improve the accuracy.

The main contributions are as follows:

- We achieve significant accuracy improvement by introducing statistical modeling of multi-sensor data.
- We reduce time accumulation effect and improve the accuracy by extracting CO_2 features with physical model.
- We show that the estimation accuracy is improved by using multiple sensor data compared to single source data.

The rest of this paper is organized as follows. Section II formulates the problem. Section III presents the transitional method: MAP estimation of HMM, MSPRT and ANN. Section IV describes feature extraction and the statistical model used in the estimation. Section V presents the experimental results on real data. Section VI concludes the paper.

II. Problem Formulation

In order to estimate the occupancy levels, we collect the CO_2 concentration and the plugs power data in the room. We adopt time-discrete model in our problem formulation and set the time granularity to one minute, as it is sufficient for meeting the demands of most practical engineering application in smart building HVAC and lighting systems.

Let y_k denote the observation state at time k consisting of the CO_2 concentration and the plugs power, and x_k denote the state of occupancy levels at time k. Our task is to infer the $\{x_k\}$ sequences from the $\{y_k\}$ sequences. The problem can be further described as the estimation of hidden states through observations, where the observation model can be written as

$$y_k = g(x_1, x_2, ..., x_k, v_k), \tag{1}$$

where v_k is observation noise. The whole process can be divided into two phases:

1) Training phase: we collect a sample observation state sequence $\{y_k\}$ and the corresponding hidden state sequence $\{x_k\}$, to establish the above observation model.
2) Estimation phase: the occupancy level is estimated by applying the model obtained during the training phase together with collected observation sequence $\{y_k\}$.

We adopt a linear model for plugs power and use the diffusion equation to model CO_2 concentration, which will be described in Section IV. If there exists relationship between the hidden state x_{k-1} and x_k as

$$x_k = f(x_{k-1}, u_{k-1}) \tag{2}$$

where u_{k-1} is system noise. This problem can be formulated as an inference problem for a HMM. On the other hand, when the hidden state x_k is independent of each other, MSPRT or ANN can be used to estimate the hidden state.

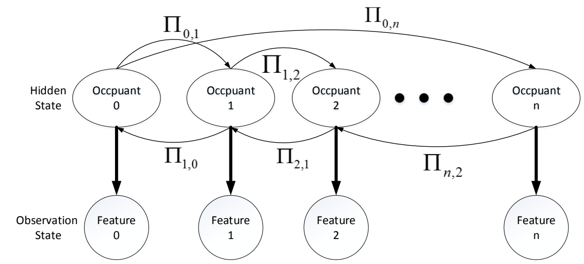

Fig. 1. Structure of Hidden Markov Model

III. Related Work

A. MAP estimation of HMM

A typical method used to estimate the hidden state is Hidden Markov Model (HMM) [13]. In this method, the occupancy level x_k is considered to be the hidden state and the feature y_k is the observation. In Hidden Markov Model (see in Fig. 1), the transition probability between previous state x_{k-1} and current state x_k is $\Pi_{x_{k-1},x_k} = \Pr(x_k|x_{k-1})$, the corresponding observation is y_k. In order to infer the occupancy level, we can perform maximum a posteriori (MAP) estimation as

$$x_k = \operatorname*{argmax}_x \Pr(x_k = x|y_1, y_2, ..., y_k) \tag{3}$$

On the premise that the state transition probability matrix and observations probability matrix are known, the above problem can be solved by the Viterbi algorithm [13]. We take $\delta_k(x) = \Pr(x_k = x|y_1, y_2, ..., y_k)$ as the likelihood function. We use $x' = \varphi_k(x)$ to store the most likely hidden state at time $k-1$ given that the hidden state at time k is x. Then we have the following relation as

$$\delta_k(x) = \max_{x'} \delta_{k-1}(x')\Pi_{x_{k-1},x_k}\Pr(y_k|x_k = x) \tag{4}$$

$$\varphi_k(x) = \operatorname*{argmax}_{x'} \delta_{k-1}(x')\Pi_{x_{k-1},x_k}\,. \tag{5}$$

Alternatively, the state estimation problem can also be solved by Bayesian filtering techniques according to the minimum mean square error (MMSE) criterion.

B. Multiple-Hypothesis Sequential Probability Ratio Test (MSPRT)

Sequential Probability Ratio Test (SPRT) is a statistical procedure to make a decision between two hypotheses [14]. There are two alternate hypothesis H_0, H_1 for a series of random variables $Y^n = (y_1, y_2, ..., y_n)$ and let $f_0(Y^n), f_1(Y^n)$ represent the probability density, then the log likelihood function is

$$L(Y^n) = \log\frac{f_1(Y^n)}{f_0(Y^n)} = L(Y^{n-1}) + \log\frac{f_1(y_n)}{f_0(y_n)}\,. \tag{6}$$

We calculate the ratio of the probabilities and the decision is made as follows: (1) if $L(Y^n) < B$, stop sampling and accept the hypothesis H_0 (2) if $L(Y^n) > A$, stop sampling and accept the hypothesis H_1 (3) if $B < L(Y^n) < A$, take another observation, where $A,\ B$ are two constants.

The occupancy estimation is a multiple-hypothesis test problem. Since the occupancy level is slow varying, we assume a short time interval where x_k keeps constant. The samples which we need to reach a decision are from the distribution of the same hypothesis. We have M hypothesis $H_j (j = 0, 1, ..., M-1)$ and the corresponding probability density function is $f_j (j = 0, 1, ..., M-1)$. Assume that the prior probabilities of the hypotheses are known, and let π_j denote the prior probability of hypothesis H_j. When n observations are available, we can obtain the posterior probability of the k^{th} hypothesis using Bayes' rule [15]

$$p_n^k = \frac{\pi_k \prod_{i=1}^{n} f_k(y_i)}{\sum_{j=0}^{M-1} \pi_j (\prod_{i=1}^{n} f_j(y_i))} \tag{7}$$

When the posterior probability of a certain hypothesis exceeds the corresponding threshold value, the decision is made to accept the hypothesis. Otherwise continue sampling the data. The M hypotheses can be used to test the states of $M-1$ occupancy levels by sequentially applying MSPRT to the observation sequences.

C. Artificial neural network

Artificial Neural Network (ANN) is a mathematical or computational model inspired by biological neural networks. It provides a universal and practical method for approximating functions from inputs. The mapping function of a neural network is defined as:

$$\begin{aligned} y^k &= \Psi''\left(\sum_j \omega_j'' h_j(x_k) + \eta''\right) \\ h_j(x_k) &= \Psi'\left(\sum_i \omega_{j,i}' f_i(x_k) + \eta'\right) \\ f_i(x_k) &= \Psi\left(\sum_k \omega_{i,k}(x_k) + \eta\right) \end{aligned} \tag{8}$$

where Ψ'', Ψ', Ψ are activation functions. The process of training an artificial neural network is to adjust the weight $\boldsymbol{\omega}$ to minimize the error between output values and target values. Once the neural network model has been obtained, it can be used to estimate the occupancy levels by computing outputs for the testing data. We implement a fully connected neural network with two hidden layers in this study and each hidden layer is composed of 8 neurons. The output layer activation function is a linear function, and the hidden layer activation function is a sigmoid function. The input vector includes CO_2 features and the total plugs power.

IV. Methodology

A. Feature extraction

Since the variation of CO_2 concentration has time accumulation effect, it cannot effectively reflect the change of the occupancy level. The estimation will be more accurate if the observation state could better reflect the occupancy level variation. We extract features with physical diffusion model

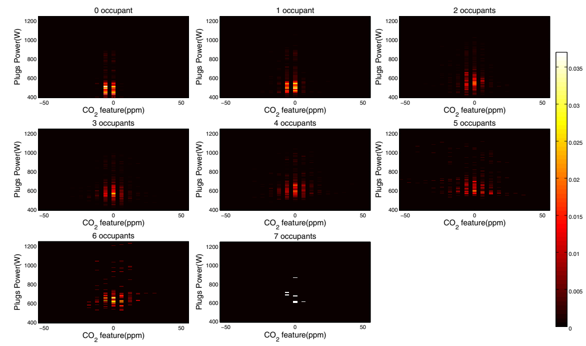

Fig. 2. Joint distribution of CO_2 features and plugs power

of CO_2 concentration [16]. The diffusion equation can be expressed as

$$V\frac{dC_R}{dt} = m_s(C_S - C_R) + S \cdot x \tag{9}$$

where V is the room volume, C_R is the indoor CO_2 concentration, C_S is the outdoor CO_2 concentration, m_S is the supply air flow rate, x is the indoor occupancy level, S is the average human emission rate per person.

The variation of CO_2 concentration is caused by outdoor air exchange and CO_2 breathed out by humans. Hence, by feature extraction, we take CO_2 emission rate of the humans as the observations. The process of feature extraction is to calculate the unknown signals $S \cdot x$. Some parameters are unknown. We can obtain the CO_2 concentration sequences and the corresponding occupancy level sequences from the training data. Outdoor CO_2 concentration C_S is modeled in a simple manner with a constant value. Thus we can obtain those unknown parameters by the least squares method:

$$\begin{aligned} \boldsymbol{p} &= (\boldsymbol{A^T A})^{-1}\boldsymbol{A^T m} \\ \boldsymbol{A} &= \begin{bmatrix} C_R(2) - C_R(1) & x(1) \\ \vdots & \vdots \\ C_R(n) - C_R(n-1) & x(n-1) \end{bmatrix} \\ \boldsymbol{p} &= \begin{bmatrix} \frac{m_s}{V} & \frac{S}{V} \end{bmatrix}, \boldsymbol{m} = \begin{bmatrix} C_S - C_R(1) \\ \vdots \\ C_S - C_R(n-1) \end{bmatrix} \end{aligned} \tag{10}$$

After obtaining the parameter $\frac{m_s}{V}$, we can obtain the observation state by

$$y(i) = C_R(i+1) - C_R(i) - \frac{m_s}{V}(C_S - C_R(i)) \tag{11}$$

where $y(i)$ is the feature that better reflects the actual CO_2 breathed out by the occupants and thus the variation of occupancy levels.

B. Statistical model

It is required to use empirical distribution of the environment parameters with different number of occupants in HMM and MSPRT. We use the joint distribution of the total plugs power in the room and CO_2 features. Due to the lack of training data, the joint distribution is discontinuous. As shown in Fig. 2, values of some points are absent. Increasing

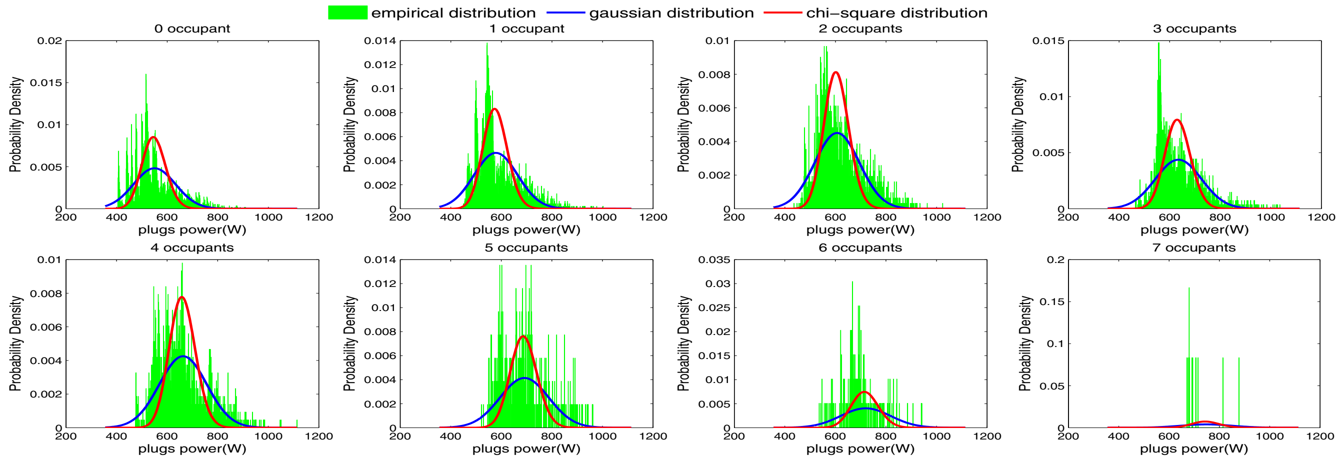

Fig. 3. Empirical and Statistical model distribution of plugs power

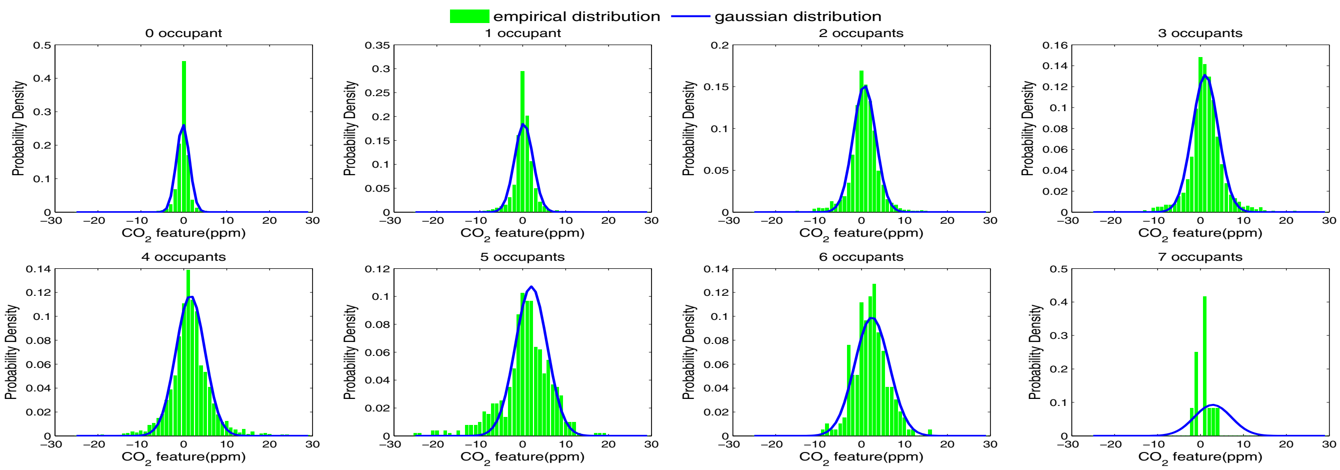

Fig. 4. Empirical and Gaussian distribution of CO_2 features

the training time can compensate for the data deficiency at the cost of extra workload. Another problem is to choose the appropriate quantization step when we use the empirical distribution. Large quantization step will lose accuracy, while small quantization step is fragile to noise perturbation.

Fitting the empirical distribution with the aid of statistical models can alleviate the problem of data deficiency, and also reduce the impact caused by the noise in the empirical distribution. Suppose that the measurements are independent conditioned on the number of occupants. We can independently fit the corresponding marginal distribution and then obtain the joint distribution.

For plugs power, its distribution for different occupancy levels can be modeled as the sum of the distribution for nobody and that for a single user. Let R denote the plugs power when no one in the room, S denote the plugs power consumed by a single user. Thus when m people are in the room, the total plugs power T_m is

$$T_m = R + \sum_{i=1}^{m} S_i, \;\; S_i \text{ are i.i.d}. \tag{12}$$

We fit the distributions of R and S with Gaussian distribution for two reasons: the bell shape of the histograms from empirical data and that the sum of Gaussian random variables remains Gaussian distribution. Suppose that R and S respectively follow the Gaussian distribution $N(\mu_r, \sigma_r^2)$ and $N(\mu_s, \sigma_s^2)$. Then T_m follows the Gaussian distribution $N(\mu_r + m\mu_s, \sigma_r^2 + m\sigma_s^2)$. By utilizing maximum likelihood estimation (MLE) on the training data, we can derive the parameters of the Gaussian distribution. We have a series of data samples $\{y_0(k)\}, \{y_1(k)\}, ..., \{y_m(k)\}$ which follow the distribution $f_R(y), f_{T_1}(y), ..., f_{T_m}(y)$ respectively. Thus the log likelihood function is

$$\begin{aligned}
&\ln[L(\mu_r, \mu_s, \sigma_r, \sigma_s)] \\
&= \sum_{i=0}^{m} \Bigg(-\frac{n_i}{2} \ln(\sigma_r^2 + i\sigma_s^2) - \frac{n_i}{2} \ln 2\pi \\
&\quad - \frac{1}{2(\sigma_r^2 + i\sigma_s^2)} \sum_{k=1}^{n_i} (y_i(k) - \mu_r - i\mu_s)^2 \Bigg)
\end{aligned} \tag{13}$$

The maximum value of log likelihood function can be obtained by gradient descent method. Once the distribution $f_{T_m}(y)$ is obtained, we can replace the empirical distribution with $f_{T_m}(y)$ in the HMM or MSPRT.

Considering the non-negativity of plugs power, we can also use Chi-Square distributions to fit the data. Suppose

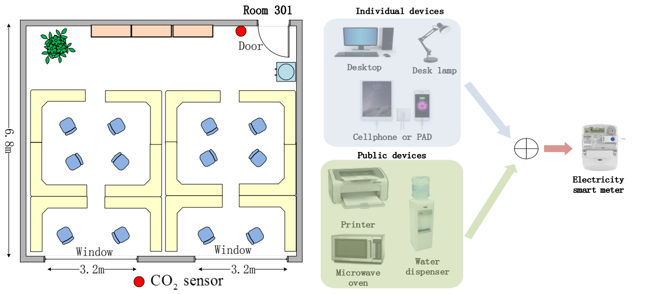

Fig. 5. Test bed layout and the measured devices of smart meters

that R and S respectively follow the Chi-Square distribution $\chi^2(n_r, \lambda_r)$ and $\chi^2(n_s, \lambda_s)$. Thus T_m follows the Chi-Square distribution $\chi^2(n_r+mn_s, \lambda_r+m\lambda_s)$, and the parameters can be obtained by MLE. The empirical distribution, Gaussian distribution and Chi-Square distribution of plugs power with different occupancy levels is shown in Fig. 3. Similarly, we fit the empirical distribution of the CO_2 features with Gaussian distribution, as shown in Fig. 4.

V. Experiment

A. Experiment setup

The experiment is performed in room 11-301, Rohm Building, Tsinghua University for two weeks from November 28th to December 12th, 2015. The room area is about $58m^2$ and the room layout is shown in Fig. 5. We collect CO_2 data by a BACnet system and collect the total plugs power data by wireless smart meters. The CO_2 sensor used in the experiment is Siemens QPA20, which is located near the door. As shown in Fig. 5, the total plugs power includes individual devices power and public devices power. Different people have different habits of using individual devices, but public devices are always kept on. The sampling period is one minute. We have installed a web camera to record occupants change in the room and obtain ground truth. The training phase is from November 29th to December 7th, and the testing phase is from December 7th to December 12th.

B. Performance indexes

We take two performance indexes into consideration: root mean square error (RMSE) and false positive (FP) rate.

RMSE refers to the average error between the estimated occupancy level and the real occupancy level, reflecting the accuracy of the algorithm. Assuming that x_k, $\hat{x}_k$ represent the ground truth occupancy level and the estimated occupancy level at time k, the RMSE is

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\hat{x}_k - x_k)^2} \tag{14}$$

FP rate is an error rate which indicates estimating the room to be occupied while it is not actually. It reflects the estimation accuracy of whether the room is occupied. Accurately judging whether the room is occupied is important to automatic control and energy saving. We introduce the set

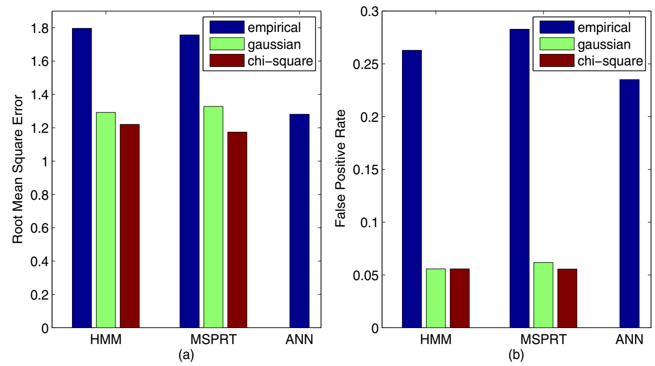

Fig. 6. The performances of different methods: (a) root mean square error and (b) false positive rate

representing the time with no one in the room $A = \{k|x_k = 0\}$, then the definition of FP rate is given as

$$FP = \frac{|\{k|\hat{x}_k > 0, k \in A\}|}{|A|} \tag{15}$$

C. Test results

We compare the test results of five methods: 1. MAP estimation of HMM based on empirical distribution, 2. MSPRT based on empirical distribution, 3. ANN, 4. MAP estimation of HMM based on statistical model distribution, 5. MSPRT based on statistical model distribution. In the experiment, there are two circumstances where the statistical model can be applied: 1. CO_2 features and plugs power both follow Gaussian distribution, 2. CO_2 features follows Gaussian distribution, and plugs power follows Chi-Square distribution.

Fig. 6 shows the results of the two cases. We can see that the estimation accuracy is largely improved when a statistical model is used. When we use the empirical distribution, there will probably be missing values of some points due to data deficiency. The estimation accuracy will be influenced by statistical noise. All the problems can be alleviated by the application of statistical model distribution. Compared with ANN, FP rate also enjoys a great improvement. The estimation result of plugs power distribution has been in a small rise when using Chi-Square distribution instead of Gaussian distribution. Compared with Gaussian distribution, Chi-Square distribution can better reflect the actual distribution of plugs power.

Fig. 7 shows the realizations of the estimation of the various methods for one day (Dec 9th). The estimated results of ANN fluctuate frequently, as each point in ANN is independent and estimation of each point is based on their own features. The environmental parameters have strong time correlation properties. And the HMM model can better deal with these temporal correlations because of the dynamic Markov properties. MSPRT makes a judgment based on sufficient data. Therefore, HMM and MSPRT can reduce the impact of noises and do not cause drastic fluctuations. Since the position of CO_2 sensor is close to the door in our experiment, the CO_2 concentration is easily influenced by

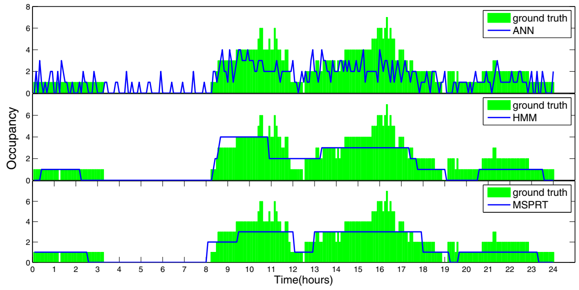

Fig. 7. Realizations of the estimation results for one day (Dec 9th) in our experiments

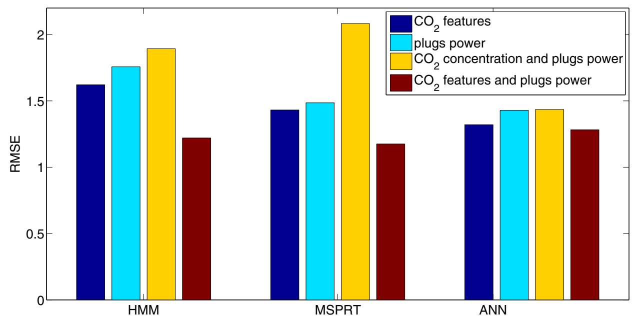

Fig. 8. Root mean square error of different data sources in experiments

the door opening and shutting, which degrades the accuracy. The frequent change of occupants also brings challenges.

As mentioned above, better features reflecting occupancy number variation would lead to more accurate estimation. Fig. 8 shows the results from original CO_2 concentration and CO_2 features extracted by diffusion model. The estimation accuracy is improved by feature extraction. The comparison result between single data source estimation and multiple data sources estimation is also shown in Fig. 8. It is obvious that the result of multiple data sources improved significantly. When signal data source estimation is adopted, an error may occur due to the interference from ambient. For instance, CO_2 concentration may change with the ventilation conditions in the room. The CO_2 concentration may drop when the door suddenly opens. Hence, estimation only based on CO_2 concentration may lead to a wrong estimation of occupancy level. Whereas, multiple data sources would not be easily disturbed at the same time, such as the change of ventilation condition will not influence the plugs power. The joint distribution value of the point is small, and hence it will not cause wrong estimation. By using multiple data sources, we can exclude the ambient disturbance and increase the accuracy of estimation.

VI. Conclusion

We propose a statistical inference scheme to estimate occupancy levels in the building rooms. Occupancy levels estimation can be formulated as a hidden state estimation problem. The environmental parameters such as CO_2 concentration and total plugs power are adopted as observations. In particular, we performed CO_2 concentration feature extraction accompanied by physical model to better reflect the change of the occupancy level. The statistical model is utilized to parameterize the empirical distribution. We verified our scheme with real experimentation. Compared with the baseline algorithms, our algorithm performs better in terms of both RMSE and FP. The use of statistical modeling can alleviate the problem of limited training data, and improve the performance of estimation. Besides, our method can be easily applied to other statistical inference methods using empirical distribution and other application scenarios. Finally, comparison between multiple data sources estimation and single data source estimation is presented, which proves that the accuracy of estimation is enhanced through multi-sensor data fusion.

References

[1] U.S. Department of Energy. Building energy data book. http://buildingsdatabook.eren.doe.gov. Accessed: 2016-02-11.

[2] European Commission. Buildings - energy. http://buildingsdatabook.eren.doe.gov. Accessed: 2016-02-11.

[3] Varick L Erickson, Yiqing Lin, Ankur Kamthe, Rohini Brahme, Amit Surana, Alberto E Cerpa, Michael D Sohn, and Satish Narayanan. Energy efficient building environment control strategies using real-time occupancy measurements. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pages 19–24. ACM, 2009.

[4] Bing Dong, Khee Poh Lam, and C Neuman. Integrated building control based on occupant behavior pattern detection and local weather forecasting. In *Twelfth International IBPSA Conference. Sydney: IBPSA Australia*, pages 14–17, 2011.

[5] Nan Li, Gulben Calis, and Burcin Becerik-Gerber. Measuring and monitoring occupancy with an rfid based system for demand-driven hvac operations. *Automation in construction*, 24:89–99, 2012.

[6] Varick L Erickson, Stefan Achleitner, and Alberto E Cerpa. Poem: Power-efficient occupancy-based energy management system. In *Information Processing in Sensor Networks (IPSN), 2013 ACM/IEEE International Conference on*, pages 203–216. IEEE, 2013.

[7] Robert H Dodier, Gregor P Henze, Dale K Tiller, and Xin Guo. Building occupancy detection through sensor belief networks. *Energy and buildings*, 38(9):1033–1043, 2006.

[8] Bing Dong, Burton Andrews, Khee Poh Lam, Michael Höynck, Rui Zhang, Yun-Shang Chiou, and Diego Benitez. An information technology enabled sustainability test-bed (itest) for occupancy detection through an environmental sensing network. *Energy and Buildings*, 42(7):1038–1046, 2010.

[9] Shengwei Wang, John Burnett, and Hoishing Chong. Experimental validation of co2-based occupancy detection for demand-controlled ventilation. *Indoor and Built Environment*, 8(6):377–391, 1999.

[10] Davide Calì, Peter Matthes, Kristian Huchtemann, Rita Streblow, and Dirk Müller. Co 2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings. *Building and Environment*, 86:39–49, 2015.

[11] Ebenezer Hailemariam, Rhys Goldstein, Ramtin Attar, and Azam Khan. Real-time occupancy detection using decision trees with multiple sensor types. In *Proceedings of the 2011 Symposium on Simulation for Architecture and Urban Design*, pages 141–148. Society for Computer Simulation International, 2011.

[12] Khee Poh Lam, Michael Höynck, Bing Dong, Burton Andrews, Yun-Shang Chiou, Rui Zhang, Diego Benitez, Joonho Choi, et al. Occupancy detection through an extensive environmental sensor network in an open-plan office building. *IBPSA Building Simulation*, 145:1452–1459, 2009.

[13] Yariv Ephraim and William JJ Roberts. Revisiting autoregressive hidden markov modeling of speech signals. *IEEE Signal processing letters*, 12(2):166–169, 2005.

[14] Peter J Bickel and Bo Li. Mathematical statistics. In *Test*. Citeseer, 1977.

[15] Carl W Baum and Venugopal V Veeravalli. A sequential procedure for multihypothesis testing. *Information Theory, IEEE Transactions on*, 40(6), 1994.

[16] Shengwei Wang and Xinqiao Jin. Co 2-based occupancy detection for on-line outdoor air flow control. *Indoor and Built Environment*, 7(3):165–181, 1998.