Real-time Building Occupancy Sensing Using Neural-Network Based Sensor Network

Tobore Ekwevugbe, Neil Brown, Vijay Pakka
Institute of Energy and Sustainable Development
De Montfort University
Leicester, United Kingdom
tekwevugbe@dmu.ac.uk

Denis Fan
Department of Energy and Climate Change
London, United Kingdom

Abstract—**Current occupancy sensing technologies may limit the effectiveness of buildings controls, due to a number of issues ranging from unreliable data, sensor drift, privacy concerns, and insufficient commissioning. More effective control of Heating, Ventilation and Air-conditioning (HVAC) systems may be possible using a smart and adaptive sensing network for occupancy detection, capable of turning off services out of hours, and not over-ventilating, thus enabling energy savings, and not under-ventilating during occupied periods, giving comfort and health benefits. A low-cost and non-intrusive sensor network was deployed in an open-plan office, combining information such as sound level, case temperature, carbon-dioxide (CO_2) and motion, to estimate occupancy numbers, while an infrared camera was implemented to establish ground truth occupancy levels. Symmetrical uncertainty analysis was used for feature selection, and a genetic based search to evaluate an optimal sensor combination. Selected multi-sensory features were fused using a neural network. From initial results, estimation accuracy reaching up to 75% for occupied periods was achieved. The proposed system offers promising opportunities for improved comfort control and energy efficiency in buildings.**

Keywords- Occupancy, Features, Sensors, HVAC systems, Energy savings

I. INTRODUCTION

Approximately about 40% of the world's energy is consumed by buildings [1], of which roughly about half of this energy is consumed by heating, ventilation, and air conditioning (HVAC) systems [2]. Significant energy is often wasted servicing unoccupied buildings: for example, [3], found 39% of US domestic building energy wasted due to unoccupied heating and cooling. There is clearly, great potential for energy savings through improved HVAC operations.

Current HVAC systems in most office buildings are operated based on fixed schedules, assuming maximum occupancy during occupied hours (typically between 9.00am and 6pm), and zero occupancy during nights and weekends. Clearly, this policy will not maximise energy savings, and does not consider periods when buildings are partially occupied. For instance, during the day, individual offices may be in use regularly while other rooms such as conference rooms may be left empty or used semi-regularly [4]. Again, convectional HVAC operations just make use of temperature and humidity as sole inputs for system control, which often leads to energy waste [5]. One possible solution for achieving energy efficiency in buildings is to couple real-time occupancy information to building controls, such that services are provided only when needed (i.e during occupied instances), whilst maintaining satisfactory indoor comfort. Previous studies have proposed up to 56% HVAC related energy savings with the application of demand-driven HVAC operations [6, 7]. A number of occupancy detection systems in the literature, have certain short-comings with respect to accuracy, cost, intrusiveness, and privacy. This paper attempts to address these limitations, by fusing information from a network of low-cost sensors for building occupancy detection. This study is distinguished from previous research in that it introduces the use of symmetrical uncertainty analysis for feature selection, and a genetic based search to evaluate an optimal sensor combination for occupancy estimation. It goes further to investigate new method of occupancy sensing: the use of case temperature, and sensing sound level as an event. To the best of the authors' knowledge, these tools have not been examined for occupancy detection.

II. RELATED WORKS

Conventional occupancy detection systems have several short-comings; Passive infrared (PIR) sensor is the most commonly used technology for occupancy sensing in non-domestic buildings especially for lighting control [8], however it fails to detect stationary occupants, thus switching off services falsely. [9] proposed a smart occupancy sensor that adapts to changing activity levels of occupants in a building zone. The authors demonstrated that by varying a PIR sensor time delay, with respect to a known activity pattern of an occupant, the number of false-off's can be minimised. However, in cases where occupancy patterns are uncertain, variation in time delay alone may not completely eliminate the problem of false-offs. To address this problem, PIR sensors are coupled with other sensors. [10] proposed a Bayesian belief network, comprising of three PIR sensors and a telephone sensor to probabilistically infer occupancy. Occupied state of individual offices room was modelled with a

Markov chain. Their system had a detection accuracy of 76%, but was unable to count the number of occupants. In an attempt to improve the robustness of occupancy numbers detection, [11-13] proposed a system that used information from carbon-dioxide (CO_2), acoustic and PIR sensors to estimate the number of occupants in an open-plan office space. Using information theory, the most relevant information for occupancy prediction was extracted from sensor data, and fused with three machine learning algorithms (support vector machine, artificial neural networks, and hidden Markov model). An average reported accuracy of 73% was achieved by the hidden Markov model.

A number of studies have highlighted the feasibility of occupancy detection in offices, by monitoring office equipment usage. For example, [14] developed a novel occupancy detection system running on an existing IT infrastructure. The system monitored occupants' MAC and IP addresses, keyboard and mouse activities as occupancy proxies. A detection accuracy of 80% at the building level, and 40% at the floor level was achieved. [15] proposed a useful method for establishing usage patterns of electrical appliances (such as desktop PCs), from which occupancy could also be inferred. Using portable temperature sensors attached to the case of PCs, and a pinging software routine that runs on the local network, appliance duty cycles were detected to a precision in excess of 97%. Both systems are unable to detect occupants not using a computer. Vision-based systems have also been used [16, 17], although occupants' privacy is a concern, besides their applicability is limited in heavily partitioned spaces. The use of wearable sensors for monitoring occupants have also been reported [18], although occupants' willingness to wear the devices may be a critical factor for its uptake, especially in office buildings.

III. INSTRUMENTING THE SPACE

The test area chosen for data gathering and system development is an admissions office within the Queen's building - this is an advanced naturally ventilated building, which forms part of the De Montfort University campus in the English Midlands, and was constructed in the early 1990s. The test area is an open-plan office which accommodates 6 members of staff, has a small kitchen capable of containing about two persons at a time. There are 6 desktop computers and a printer in the space. The room has one exit door, high ceiling and a large rear window that is usually kept locked, although there are smaller windows at the side which are often put to use by occupants. Being an admissions office, both staff and students frequently enter the space to make enquiries, indoor comfort conditions can quickly deteriorate as a result of indoor variables build-up. At best, the indoor climate can be considered as heavily dynamic. Fig. 1 gives an illustration of sensors deployed for indoor environmental monitoring.

Various technologies were used to gather data, HOBO temperature sensors were attached to the case of all six desktop computers to infer usage pattern, while self contained HOBO U series dataloggers were employed to monitor the indoor climate, including temperature, humidity, and illumination. Volatile Organic Compounds (VOC's) and CO_2 levels were monitored using Aerasgard rlq-series air quality sensor, and four GE Sensing Telaire CO_2 sensors respectively, results being logged using HOBO dataloggers. Ambient sound levels were monitored using custom designed circuitry, which would record as an event sound level over a preset threshold. Traditional PIR sensors (three in number), as used for lighting control, were installed to monitor motion. All events were recorded using HOBO event loggers. Though sensor measurements were acquired every one minute, a five minute average was used for analysis. Data were downloaded to a PC using a HOBO shuttle and uploaded to a *MySQL* database using *MATLAB* scripts. Occupancy estimates from the data were subsequently analysed in *MATLAB,* and Waikato Environment for Knowledge Analysis (*WEKA*) [19].

An infrared camera was mounted in the test area to capture occupants' traffic. Video capture and recording was done using an ordinary laptop, with images captured at a one-minute interval. This was also found to be a more efficient approach than, capturing live streaming video, with no significant loss of resolution. Occupancy numbers validation was carried out by manually counting the number of occupants in each image. This information is referred to as the ground truth occupancy numbers in this work, and was used for model training and testing.

Sensor data were collected between 12/09/2012 00:00:00am and 11/10/2012 23:55:00pm. All corrupted and missing data instances due to instrumentation limitations (e.g need to charge batteries used to power CO_2 sensors, or laptop may need restarting after image feeds become static) were excluded. Overall, after the pre-processing stage, 10000 data instances were used for analysis, 8000 for model training, and 2000 for testing.

IV. METHODOLOGY

A central objective of this study is to arrive at the combination of environmental ambient sensors that provides the most relevant information for detection of occupancy numbers in an observed environment. This entails the use of feature selection algorithms, which can be broadly classified into two categories: The filter and wrapper model. The filter model relies on general characteristics of the training data to select features without involving any learning algorithm, they are computationally cheap, and do not inherit any bias of a learning algorithm [20]. On the other hand, the wrapper model uses the predictive accuracy of a predetermined learning algorithm to search the feature space, in order to determine features subsets with the highest quality. It tends to be more computationally expensive and time consuming than the filter model, and thus may not be practical to apply for feature selection when the data set is large containing numerous features and instances [21]. Information theory (IT) falls in to the filter model category, and it is a widely used non-linear correlation measure for feature relevance analysis [22], besides it has been used in the development of occupancy detection system [11]. Here, the features obtained from individual sensors in the sensing network are explored for occupancy detection, using information theory based analysis. A brief overview of the feature selection process is presented.

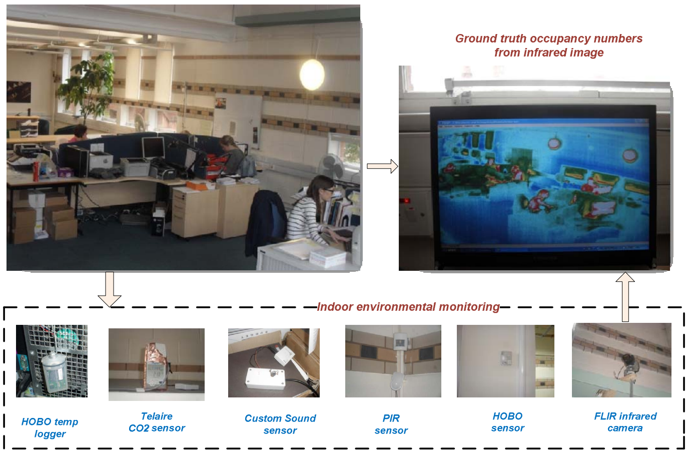

Figure 1. Test area instrumentation

A. *Information Theory*

In information theory, Entropy is a measure of the amount of uncertainty of a random variable. The entropy of a variable Y is defined as in (1), and the entropy of Y before and after observing values of another variable X is given by (2).

$$H(Y) = -\sum_i p(y_i)\log_2 p(y_i) \qquad (1)$$

$$H(Y|X) = -\sum_j p(x_j)\sum_i p(y_i|x_j)\log_2 p(y_i|x_j) \qquad (2)$$

Where $p(y_i)$ is the prior probability for all i^{th} values of Y and $p(y_i|x_j)$ is the conditional probability of y_i given x_j.

Suppose Y and X represents the set of classes and features respectively in a given data set. Entropy is 0 without any uncertainty if all subsets of a feature belong to the same class. On the other hand, subsets in a feature space are totally random to a class if entropy is 1. The maximum value of entropy is 1. The amount by which the entropy of Y decreases reflects additional information about Y provided by X, and is called information gain as shown in (3).

Information gain measures the dependence between the feature and the class label, and it is defined as:

$$I(Y,X) = H(Y) - H(Y|X) \qquad (3)$$

Information gain is symmetrical for two random variables X and Y; this property is desirable for measuring correlations between features. However, it is biased for features with more values. Therefore, in this work, symmetric uncertainty (SU) is employed to determine the predictive strength of features investigated for occupancy numbers estimation. SU compensates for information gain's bias towards features with more values, and normalizes its values from 0 to 1 to ensure they are comparable. It treats pairs of features symmetrically, averages the values of two random variables, hence it has no bias problem. This methodology has been proven to be efficient in determining relevant features in many machine learning applications, capable of improving a classifier's accuracy [20].

$$SU(Y,X) = 2\left[\frac{I(Y,X)}{H(X)+H(Y)}\right] \qquad (4)$$

If the symmetric uncertainty evaluation measure of a feature to the class label is low, it implies the feature has poor predictive ability to the class, and vice-versa. Features can be ranked in descending order according to their degrees of association to the class label Y such that $SU(Y, X_i) \geq SU(Y, X_j)$ where X_i and X_j are two features.

B. *Feature Selection*

Raw sensor measurements were initially subjected to pre-processing, which included timestamp synchronization, removal of outliers, and missing values. New sets of features were then created based on the pre-processed sensor data. These features are intended to capture temporal variations in indoor climatic measurements. Features analysed, alongside their description is given in Table I.

TABLE I. FEATURES INVESTIGATED AND THEIR DESCRIPTIONS

Air temperature, relative humidity, VOC and CO_2 measurements	
First order difference (FIR)	raw (i) - raw (i) -1
Second order difference(SEC_FIR)	raw_FIR (i) - raw_FIR (i) -1
5 minute moving average (AVR_5min)	$((\sum_{i-4}^{i} raw(i))/5)$
Approximate area under the curve for between two data instances (AREA)	$\int_{t(i)-1}^{t(i)} FIR(i) \approx FIR(i) - FIR(i-1)(f(FIR(i)) + f(FIR(i-1))/2$
Sound and PIR data	
Total number of times sound or motion is detected in a minute (PUL_1min)	
5 minute moving average of sound or motion detected (PUL_5min)	$((\sum_{i-4}^{i} PUL\_1min(i))/5)$
Occupied times : Duration of occupancy as detected by the sound or PIR sensor in a minute (OCC_1min)	
Occupied times : Duration of occupancy as detected by the sound or PIR sensor in a minute (OCC_5min)	$(\sum_{i-4}^{i} OCC\_1min(i))/5)$
Case temperature measurement	
Case temperature of desktop computer recorded after 1 minute (CAS_1min)	
5 minute moving average of CAS_1min (CAS_5min)	$(\sum_{i-4}^{i} CAS\_1min(i))/5)$

TABLE II. SENSOR FEATURES RANKING ACCORDING TO SU MEASURE VALUES

Features	**Features description**	**SU measure**
AVR_temp_5min	Average ambient temperature measurement in 5 minutes interval	0.0643
CAS_5min	Average case temperature measurement in 5 minutes interval	0.2542
PUL_SND_5min	Number of times the sound sensor is activated in 5 minutes interval	0.1974
AREA_RH_5min	Area under the curve for relative humidity levels in 5 minutes interval	0.0829
AREA_VOC_5min	Area under the curve for VOC levels in 5 minutes interval	0.0985
PUL_PIR_5min	Number of times the PIR sensor is activated in 5 minutes interval	0.1687
AREA_CO_2_5min	Area under the curve for CO_2 levels in 5 minutes interval	0.2940

The predictive ability of features for each individual sensing domain was determined using SU. Features with the highest predictive value in each were passed on to a genetic based Correlation Feature Selection (CFS) filter, to determine an optimal combination of features for occupancy detection, since SU and most feature weighting/ ranking algorithms are incapable of removing redundant features because redundant features are likely to have similar rankings or predictive power [22]. Table (II) gives the features with the highest SU measure values for each sensing domain investigated. CFS uses a correlation based heuristics called "merit" to evaluate the worth of features.

$$Merit_S = \frac{k\overline{r_{cf}}}{\sqrt{k+k(k-1)\overline{r_{ff}}}} \quad (5)$$

$Merit_S$ is the heuristic "merit" of a feature subset S containing k features, and $\overline{r_{cf}} = \sum_{f_i} \epsilon S \frac{1}{k} \sum(f_i, C)$ is the mean feature class correlation and $\overline{r_{ff}}$ is the average feature inter-correlation. CFS explores the feature space for an optimal combination of a feature subset with the highest $Merit_S$ value using a genetic algorithm based search, with a stopping criterion set at when this value does not increase. The feature selection process is carried out in *WEKA* [19]. Results indicated that CO_2, PIR, sound and case temperature sensors provided the best combination of features for the prediction task. A detailed description of CFS algorithm can be seen in [23].

C. Neural Network Based Fusion for Occupancy Detection

A Feed-Forward Neural Network was applied in this study, for model training, testing and validation. Artificial Neural Networks (ANN's) are biologically inspired systems used for model estimation in which a set of variables are estimated through training from available data. The neural net system comprises of a set of input and output variables used for learning the model responsible for this data. These networks consist of a number of individual units called neurons. Connections between neurons have certain weights that are usually obtained using some learning rules. A neural net of two hidden layers with same combination of neuron numbers in each hidden layer was tested on indoor climatic data from the test area. The learning algorithm employed is the back-propagation algorithm, where the network error is back propagated from the output to input layer. Within the network, data are subjected to simple processing within its layers, and the weights of each neuron are adjusted in order to minimize the mean-squared error between the input and the target data, according to a specified accuracy index, or after the completion of a specified number of iterative learning processes. Once the ANN model has been satisfactorily trained, and tested, it is used to predict output data from previously unseen input data. In this work, ANN was implemented using the *MATLAB* Neural Network toolbox. An optimal combination of features resulting from the feature selection process described in section *B* (AREA_CO_2_5min, PUL_SND_5min, CAS_5min, PUL_PIR_5min) were used as inputs for the ANN. The Log Sigmoid transfer function was used in both hidden layers, while a linear function was used in the output layer. 15 neurons were used in each of the connecting layers, with other parameters such as a learning rate of 0.05, number of epochs of 500 and momentum of 0.9. The training phase is repeated for 10 times to increase the probability of reaching a global solution. The resulting average from the outputs of the training phases is used for analysis in next stages.

V. RESULTS AND DISCUSSION

The test results were evaluated to provide occupancy information, such that HVAC systems can be proactively adjusted based on it. In this work, occupancy detection does not indicate the rate of sensing occupancy presence (e.g occupied and unoccupied), instead it refers to the rate of sensing occupancy numbers in the test area.

Two standard statistical performance evaluation measures, i.e Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE) are employed to validate the model performance. RMSE measures the difference between predicted occupancy and actual occupancy, while MAPE which gives the model accuracy, is used to make a term-by-term comparison of the relative error in the model predictions with respect to ground truth occupancy. It makes more sense to apply MAPE only to measurements during the occupied periods than otherwise.

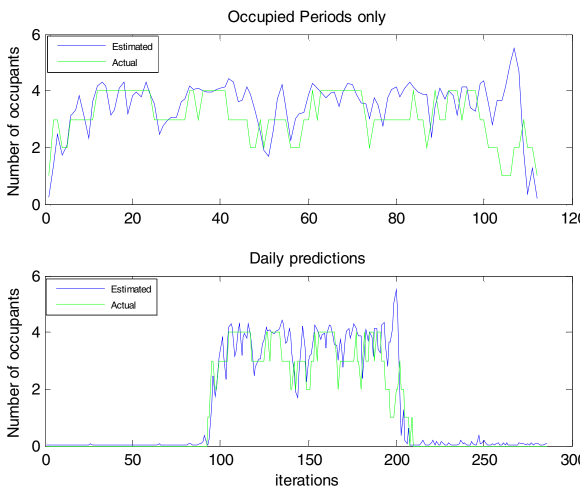

Figure 2. Occupancy estimation for 08/10/2012

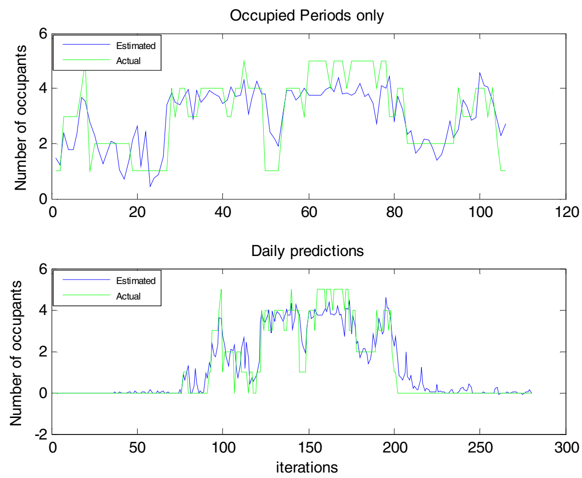

Figure 3. Occupancy estimation for 11/10/2012

Fig. 2 and 3 show model predictions and actual occupancy data for two typical days. It is clear from these plots that model predictions are in good sync with actual occupancy numbers, with an accuracy of 67% and RMSE of 1.01 on the 08/10/2012, and 69% and RMSE 0.91 on the 11/10/2012, during occupied periods. The model particularly performs well, during unoccupied period; this is not surprising as measured indoor climatic variables do not show any significant temporal variation. Although, for 11/10/2012 during unoccupied period, model predictions indicated there were occupants in the space. This may be due to the slow CO_2 decay rates in the test area, which can sometimes take till about 8:00am the next morning for CO_2 levels to completely decay.

During occupied periods, model predictions show close tracking with ground truth data. Although, model outputs are in decimal formats, and may not represent practical observations i.e, number of occupants cannot be 4.12, hence model outputs may require quantisation. However, the outputs are still useful for occupancy driven HVAC systems, since certain level of error is acceptable. Besides, HVAC systems do not need to be very sensitive, such that it responds to slight changes in occupancy numbers. For instance, a change in the number of occupants by one, normally should not cause any significant HVAC operation, unless the space switches from occupied to unoccupied, and vice versa.

Overall, the model sometimes struggles when there are abrupt changes in occupancy levels, which again may be linked to the slow CO_2 decay rates. In addition, CO_2 sensors are slow to detect incoming occupants. For occupancy driven HVAC control operations, this may not have any significant ramification, as the system is not expected to produce a control action for abrupt occupancy changes, or short occupancy durations. From the results as shown in Fig. 4, the model achieves an average daily testing RMSE of around 1.08 for occupied periods. This is considered good, since the number of occupants varies between 0 and 6. This suggests that the model predictions are usually within 1 of the actual occupancy number. Average daily testing RMSE for unoccupied periods is around 0.13 suggesting that the model is effective for detecting unoccupied periods. Fig. 5 shows the testing results accuracy for different days in a typical week, between 02/10/2012 and 11/10/2012. Different accuracies for days of the week were

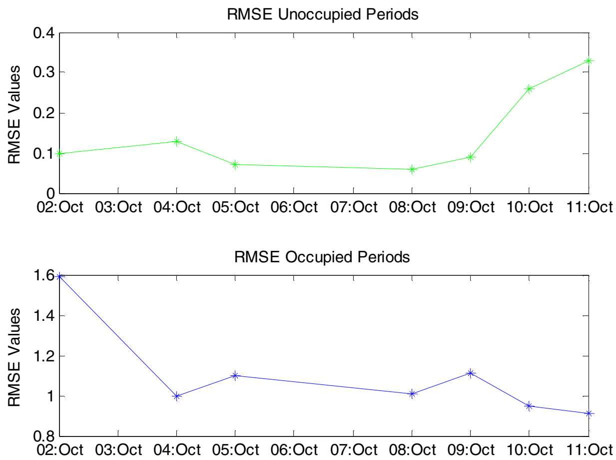

Figure 4. Model estimation RMSE from 02/10-11/10

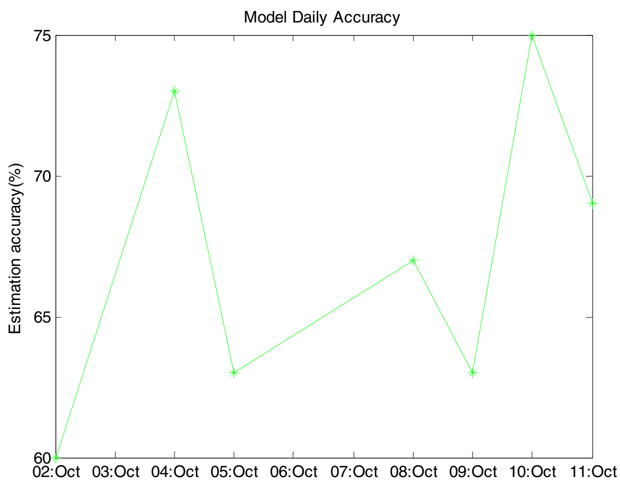

Figure 5. Model accuracy for occupied periods from 02/10-11/10

recorded, with variation reaching up to 25%. Test accuracy slightly improves to 75%, although, there are certain days where accuracy was relatively low (e.g 60% on the 02/10/2012). In summary, the model shows reasonable tracking with ground truth data during occupied and unoccupied periods.

VI. CONCLUSION

In this study, a novel methodology for estimation of occupancy numbers using symmetrical uncertainty analysis for feature selection, and a genetic based search for optimal sensor combination have been presented. New method of occupancy sensing such as the use of appliance case temperature has been introduced. Results indicate that features from CO_2, sound, case temperature, and PIR sensors have the largest correlation with the number of occupants in the test area. The prediction model employs a neural network to fuse selected sensor features for occupancy estimation. Predicted occupancy show reasonable tracking with actual occupancy during occupied period, although results are more impressive during unoccupied periods. Results may be limited to the specific environment used in the study. However, the proposed methodology has the potential for wider applicability in other building environments, subject to further testing. Future work could include further analysis to improve system accuracy for occupied periods, generalization of learned models for various environments (i.e. different rooms in a building, or entirely different buildings), testing this methodology with other standard machine learning approaches and exploring the use of occupancy estimates from the model for HVAC control applications.

REFERENCES

[1] B. Metz, O.R. Davidson, P.R. Bosch, R. Dave, and L.A. Mayer, " Contribution of working group III to the fourth assessment report of the intergovernmental panel on climate change," in Climate change 2007, Cambridge University Press, United Kingdom and New York, NY, USA, 2007.

[2] L. Pérez-Lombarda, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and Buildings*, vol. 40, pp. 394-398, 2008.

[3] R.J. Meyers, E.D. Williams, and H.S. Matthews, "Scoping the potential of monitoring and control technologies to reduce energy use in homes," *Energy and Buildings*, vol.42, pp. 563-569, 2010.

[4] V.L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, and A.E. Cerpa, "Energy efficient building environment control strategies using real-time occupancy measurements," *in Proceedings of the 1st ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, Berkeley, CA, USA, November, 2009.

[5] Y.B. Agarwal, R. Gupta, J. Lyles, M. Wei, and T. Weng, " Occupancy-driven energy management for smart building automation," *in Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, Zurich, Switzerland, November, 2010.

[6] Z. Sun, S. Wang, and Z. Ma, "In-situ implementation and validation of a CO_2-based adaptive demand-controlled ventilation strategy in a multi-zone office building," *Building and Environment*, vol. 46, pp. 124-133, 2011.

[7] Y. Tachwali, H. Refai, and J.E. Fagan, "Minimizing HVAC energy consumption using a wireless sensor network," *in Industrial Electronics Society, 2007, IECON 2007, 33rd Annual Conference of the* IEEE, Taipei, Taiwan, November, 2007.

[8] D.T. Delaney, G.M. O'Hare, and A.G Ruzzelli, " Evaluation of energy-efficiency in lighting systems using sensor networks," *in Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy Efficiency in Buildings,*Berkeley,CA, USA, November, 2009.

[9] V. Garg and N.K. Bansal, "Smart occupancy sensors to reduce energy consumption," *Energy and Buildings*, vol. 32, pp. 81-87, 2000.

[10] R.H. Dodier, G.P. Henze, D.K. Tiller, and X. Guo, " Building occupancy detection through sensor belief networks," *Energy and Buildings,* vol. 38 pp. 1033-1043, 2006.

[11] B. Dong, B. Andrews, K.P. Lam, M. Höynck, R. Zhang, Y. Chiou, and D. Benitez, "An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network," *Energy and Buildings*, vol. 42, pp. 1038-1046, 2010.

[12] K.P. Lam, M. Höynck, R. Zhang, B. Andrews, Y. Chiou, B. Dong, and D. Benitez," Information-theoretic environmental features selection for occupancy detection in open offices," *in Proceedings of the Eleventh International IBPSA Conference,* Glasgow, Scotland, 2009.

[13] K.P. Lam, M. Höynck, B. Dong, B. Andrews, Y. Chiou, and R. Zhang," Occupancy detection through an extensive environmental sensork network in an open-plan office building," *in Proceedings of the Eleventh International IBPSA conference*. Glasgow, Scotland, 2009.

[14] R. Melfi, B. Rosenblum, B. Nordman, and K. Christensen, "Measuring building occupancy using existing network infrastructure," *in Proc., 2011 International Green Computing Conference and Workshops,IEEE,* Orlando, FL, USA, July, 2011.

[15] N. Brown, R. Bull, F. Faruk, and T. Ekwevugbe, "Novel instrumentation for monitoring after-hours electricity consumption of electrical equipment, and some potential savings from a switch-off campaign," *Energy and Buildings*, vol. 47, pp. 74-83, 2011.

[16] Y. Benezeth, H. Laurent, B. Emile, and C. Rosenberger," Towards a sensor for detecting human presence and characterizing activity," *Energy and Buildings*, vol. 43, pp. 305-314, 2011.

[17] R. Tomastik, L. Yiqing, and A. Banaszuk,"Video-based estimation of building occupancy during emergency egress," *in American Control Conference,* Seattle, USA, June, 2008.

[18] I. Korhonen, J. Parkka, and M. Van Gils, "Health monitoring in the home of the future," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 22, pp. 66-73, 2003.

[19] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, "The WEKA Mining Software: An Update," 2009.

[20] S. Senthamarai Kannan, and N. Ramaraj,"A novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm," *Knowledge-Based Systems*, vol. **23**, pp. 580-585, 2010.

[21] J. Biesiada, and W. Duch, "Feature Selection for High-Dimensional Data: A Kolmogorov-Smirnov Correlation-Based Filter Solution," *in Proceedings of the 4th International Conference on Computer Recognition Systems, 2005,* Rydzyna Castle, Poland, May, 2005.

[22] L. Yu, H. Liu, and I. Guyon,"Efficient feature selection via analysis of relevance and redundancy," *Journal of Machine Learning Research*, vol. 5, pp. 1205 - 1224.

[23] M. Hall, Correlation based feature selection for machine learning, Doctoral dissertation, Dept. of Computer Science, University of Waikato: New-zealand, 1999.