REAL-TIME BUILDING OCCUPANCY SENSING FOR SUPPORTING DEMAND DRIVEN HVAC OPERATIONS

Tobore Ekwevugbe
Research Student

Neil Brown
Senior Lecturer

Vijay Pakka
Senior Research Fellow

Institute of Energy and Sustainable Development
De Montfort University
Leicester, United Kingdom
tekwevugbe@dmu.ac.uk

ABSTRACT

Accurate knowledge of localised and real-time occupancy numbers can have compelling control applications for Heating, Ventilation and Air-conditioning (HVAC) systems. However, a precise and reliable measurement of occupancy still remains difficult. Existing technologies are plagued with a number of issues ranging from unreliable data, maintaining privacy and sensor drift. More effective control of HVAC systems may be possible using a smart sensing network for occupancy detection. A low-cost and non-intrusive sensor network is deployed in an open-plan office, combining information such as sound level and motion, to estimate occupancy numbers, while an infrared camera is implemented to establish ground truth occupancy levels. Symmetrical uncertainty analysis is used for feature selection, and selected multi-sensory features are fused using a neural-network model, with occupancy estimation accuracy reaching up to 84.59%. The proposed system offers promising opportunities for reliable occupancy sensing, capable of supporting demand driven HVAC operations.

INTRODUCTION

Global warming is one of the most disturbing concerns facing humanity today due to the accelerated release of carbon dioxide (CO_2) and other greenhouse gases into the atmosphere as a result of human activities. The problem is compounded by decreasing availability of fossil fuels, increasing population, environmental and economic concerns regarding energy use. These all constitute drivers for the adoption of more sustainable ways of securing our energy needs (Shuai et al., 2011). Approximately about 40% of the world’s energy is consumed by buildings (ASHRAE, 2007), of which roughly about half of this energy is consumed by Heating, Ventilation, and Air conditioning (HVAC) systems (Pérez-Lombard et al., 2008). Reductions in HVAC related energy will go a long way in contributing to efforts aimed at delivering sustainable building energy use. Previous research have proposed up to 56% HVAC related energy savings with improvements in operation and management of HVAC systems (Sun et al., 2011, Tachwali et al., 2007).

Real- time building occupancy sensing is useful for efficient control of building services such as lighting and ventilation, enabling energy savings, whilst maintaining a comfortable environment. Occupancy information can be used for determination of HVAC heat loads (Chenda and Barooah, 2010), system running time, required heating, cooling and distribution of conditioned air, and optimal selection of temperature set points (Li et al., 2012). Ideally, building controls should automatically respond to dynamic occupancy loads. However, current building energy management system (BEMS) often lacks this capacity, as such they usually rely on fixed assumptions (such as peak occupancy loads as opposed to the optimal) to operate HVAC and electrical systems, leading to possible energy waste (Erickson et al., 2011). 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), and to optimize HVAC operations such that the flow rate of conditioned air into a space is adjusted based on optimal occupancy numbers.

Many occupancy detection systems in the literature have certain drawbacks with respect to accuracy, cost, intrusiveness, and privacy. This study 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 a new method of occupancy sensing: the use of case temperature. To the best of the authors' knowledge, these tools have not been examined for occupancy detection.

BACKGROUND

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 (Delaney et al., 2009), however it fails to detect stationary occupants, thus switching off services falsely. To address this problem, PIR sensors are coupled with other sensors. For instance, Padmanabh et al. (2009) used a combination of microphones and PIR sensors to gather occupancy information for efficient scheduling of a conference room. The room was considered occupied (meeting ongoing) if a microphone value exceeded a threshold twice in a 5-minute interval; otherwise it is classified as unoccupied (no meeting). Jianfeng et al. (2005) built a novel bathroom activity recognition system, consisting of microphone and PIR sensors. The system detected real-time sound events with a hidden Markov model to an accuracy of 87%. Both systems highlight the usefulness of sound sensing for activity monitoring, and hence occupancy presence detection. However, system functionality is prone to external interference which may limit their performance. Dodier et al. (2006) 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, Dong et al. proposed a system that used information from CO_2, acoustic and PIR sensors to estimate the number of occupants in an open-plan office space (Dong et al., 2010, Lam et al., 2009b, Lam et al., 2009a). 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. Meyn et al. (2009) improved occupancy detection accuracy, by using a sensor network comprising CO_2 sensors, digital video cameras and PIR detectors as well as historic building utilization data for occupancy estimation at the building level. The system used a receding-horizon convex optimization algorithm to infer occupancy numbers. Their system detection accuracy reached 89%. However, it was not able to estimate occupancy numbers at the room level.

Other studies have highlighted the feasibility of occupancy detection in offices by monitoring office equipment usage. For example, Brown et al. (2011) proposed a useful method for establishing the usage pattern of electrical appliances (such as desktop PCs), from which occupancy can 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%. Martani et al. (2012) studied the relationship between building occupancy and energy consumption, using number of *WiFi* connections as proxies for occupancy estimation. Overall, at the building level, occupancy accounted for between 63% and 69% variation of the total electricity consumption. Both systems were unable to detect occupants not using a computer.

The use of wearable sensors for monitoring occupants have also been reported; Li et al. (2012) proposed an occupancy detection system based on RFID tags, which reported real-time occupancy numbers and their localized thermal zones in an office building. A K-nearest neighbour algorithm was used for occupancy tracking to average zone detection accuracy of 88% for stationary occupants and 62% for moving occupants. The system offers promising potentials for occupancy monitoring. However, occupants’ willingness to wear the devices may be a critical factor for its uptake, especially in office buildings. Vision-based systems have also been used (Benezeth et al., 2011, Tomastik et al., 2008), although occupants' privacy is a concern. Besides, their applicability is limited in heavily partitioned spaces.

EXPERIMENTAL SET-UP

The test area is a space occupied by the Institute of Energy and Sustainable Development (IESD) 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 provides accommodation for 18 fulltime PhD research students and 21 staff members. It comprises of an open floor area, a kitchen, printing bay, equipment room, MSc area and 4 office rooms. About 40 desktop computers and 5 printers are placed at different locations inside the space. The open plan section of the IESD space has a high ceiling, and there is a hallway between cubicles accommodating PhD students and some staff members. Figure 1 provides a clearer description of the space.

The area enjoys good natural lighting due to its large side windows, although it is shaded from the direct effect of the sun by an adjacent part of the building. It is ventilated with three glazed roof vents. The vents are controlled together by room temperature sensors. The stack or roof vents opens up to introduce natural ventilation to cool the space. If the internal temperature is 4°C greater than the set point, then the stack dampers are opened. If the wind speed exceeds 20 mph then the stack dampers/motorised windows are closed. Heating in the Queens building is provided by three gas-fired boilers, one condensing boiler and two high frequency boilers. However, the IESD area is heated by radiators fed from one pipe loop from the group of three boilers used.

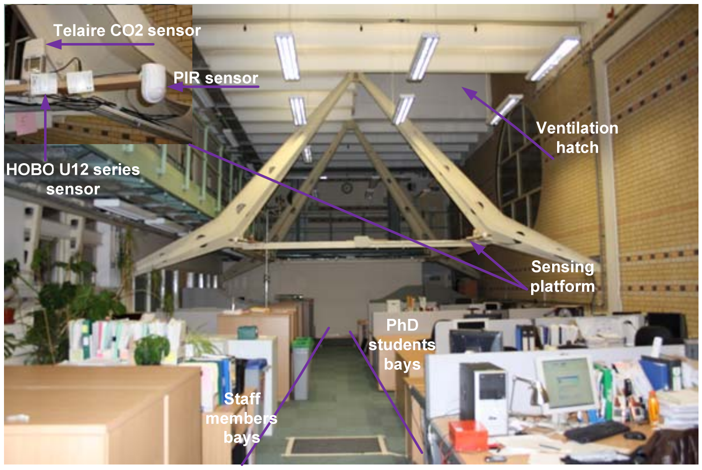

Figure 1. IESD Open plan space

In an attempt to ensure thoroughness of the monitoring campaign, the main floor was divided into several zones, which do not have any physical boundaries, and therefore, environmental conditions in the zones could interfere with one another, see figure 2. The MSc area in the open plan section of the space has been excluded from the campaign as it is rarely occupied. This test area has been chosen because of its multi-occupancy nature, and indoor environmental conditions can be considered as heavily dynamic. In addition, it provides a good representation of various components and activities within a typical university office building, and is also expected to possess capacity for good degree of experimental control.
In figure 2, six sensing platforms were deployed, with each consisting of sensors monitoring indoor environmental parameters such as temperature, lighting, relative humidity, CO_2, and motion. One sensing platform has been placed at the four corners of each zone in order to increase the sensitivity of PIR sensors, a similar approach has been utilized for monitoring occupancy pattern and health status of elderly people (Kaushik and Celler, 2007). Temperature sensors were attached to the case of desktop computers in the space, while custom-made sound sensors (which would record as an event sound level over a pre-set threshold) were placed close to occupants as depicted in figure 2.

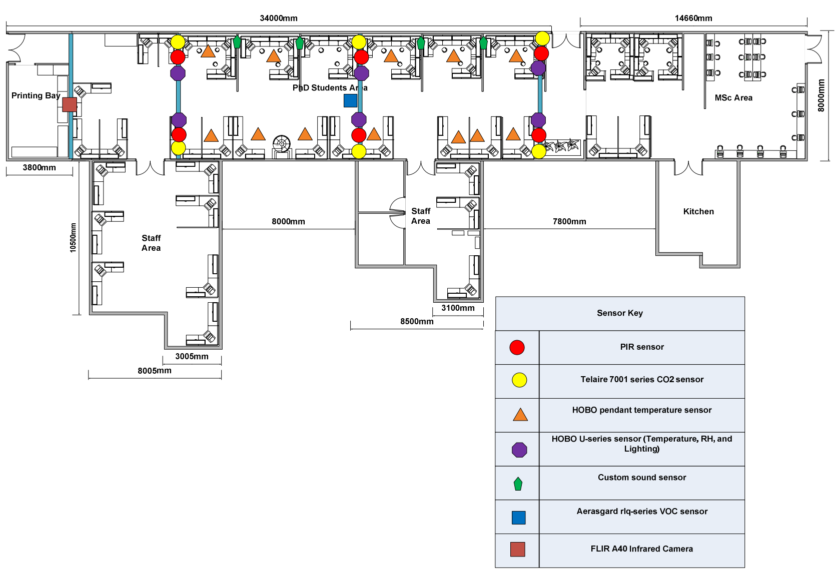

Figure 2. Sensor placement in the test area

DATA COLLECTION

Sensor data were collected for seven days between 08/07/2012 00:00:00am and 16/07/2012 10:45:00am. Though sensor measurements were acquired every one minute, a five minute average was used for analysis. Table 1 lists the raw sensor data types collected during the monitoring period. Data were downloaded to a PC, and then uploaded to a *MySQL* database using *MATLAB* scripts. Occupancy estimates from the data were subsequently analysed in *MATLAB* and Waikato Environment for Knowledge Analysis in *WEKA* (Hall et al., 2009).

An FLIR A40 infrared camera was mounted in the test area to capture occupants' traffic, see figure 3. 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 the image. This information is referred to as the ground truth occupancy numbers in this study, and was used for model training and testing. 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 from any further analysis.

Figure 3. Ground truth occupancy data using infrared camera

Table 1: Raw sensor data collected during the monitoring period

Data Type	Total data instances
CO_2 (x6)	40208
Sound (x4)	35260
Ambient RH (x6)	54720
Ambient temperature (x6)	54720
Desktop PC case temperature (x12)	97163
Motion (x6)	74220

DATA PROCESSING

Feature selection is frequently used as a data processing step in the development of many multi-sensor fusion systems. This process entails choosing a subset of features according to a certain evaluation criteria, such that the feature space is optimally reduced. It has been proven to be an effective step in removing irrelevant and redundant features, reducing data processing time, increasing efficiency in learning tasks, improving learning performance like predictive accuracy, and enhancing comprehensibility of learned results (Blum and Langley, 1997), (John et al., 1994). In this study, an information theoretical concept was applied for feature selection, in order to arrive at the combination of multi-sensory features that provided the most relevant information for detection of occupancy numbers within the observed environment. Information theory is a widely used non-linear correlation measure for feature relevance analysis in machine learning applications (Yu et al., 2004). It tends to be more computationally efficient and less time consuming than other feature selection models such as wrapper filters (Blum and Langley, 1997),(John et al., 1994). A brief overview of information theory concept is presented.

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).

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

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

Where $p(y_i)$ is the prior probability for all ith 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:

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

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) was 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 that they are comparable. It treats pairs of features symmetrically and 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, thus capable of improving a classifier's accuracy (Senthamarai Kannan, 2010).

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

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.

Feature selection process

Initial pre-processing of raw sensor measurements included timestamp synchronization, removal of outliers, and handling missing values. Thereafter, new sets of features were extracted from the pre-processed sensor data. These features were intended to capture temporal variations from indoor climatic measurements. Features analysed alongside their mathematical description are given in Table 2. A rank of extracted features for each individual sensing domain was carried out using symmetrical uncertainty analysis, in order to investigate the correlation between each feature and the number of occupants.

The first six features from each individual sensing domain as presented in table 3, were passed on successively to a genetic based Correlation Feature Selection (CFS) filter. This was used to determine an optimal combination of features for occupancy detection, since SU as with most feature weighting/ ranking algorithms are incapable of removing redundant features because redundant features are likely to have similar rankings or predictive power (Yu et al., 2004). 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 was carried out in *WEKA* (Hall et al., 2009) .

Let n Rule represent a combination that has n features in it. The $Merit_S$ score for each heterogeneous multi-sensory combination was computed using CFS, where n Rule =1, 2, 3…6. Table 4 shows the features selected for each successive combination. The most effective features were evaluated as those selected with the highest number of times, appearing over 80% (or five of the six iterations) throughout the n Rule combination process. These were considered as the best set of features for occupancy number detection within the observed environment and form the inputs for a machine learning model used for actual fusion process.

Table 2: Mathematical description of features investigated

Air temperature, relative humidity, case temperature and CO_2 measurements	
First order difference (FDIFF)	raw (i) - raw (i) -1
Second order difference(SDIFF)	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 (A)	$\int_{t(i)-1}^{t(i)} FIR(i) \approx FIR(i) - FIR(i-1)(f(FIR(i)) + f(FIR(i-1))/2$
Variance (V) Where N= Number of sensors, and $\mu =$ mean of $(AVR\_5min)_i$	$\sqrt{\frac{1}{N}\sum_{i}^{N}(AVR\_5\text{min})_i}$-$\mu_i$
Sound and PIR data	
Occupied times : Total duration of occupancy as detected by the sound or PIR sensor in 5 minutes (TOS_5min)	$\sum_{i}^{N}(AVR\_5min)_i$
Number of pulses : Total number of state changes as detected by the sound or PIR sensor in 5 minutes (TONP_5min)	$\sum_{i}^{N}(AVR\_5min)_i$
High to Low : Total number of times stages changes as from high to low as detected by the sound or PIR sensor in 5 minutes (THI_LO_5min)	$\sum_{i}^{N}(AVR\_5min)_i$
Low to High : Total number of times stages changes as from high to low as detected by the sound or PIR sensor in 5 minutes (TLO_HI_5min)	$\sum_{i}^{N}(AVR\_5min)_i$

Table 3: First six individual sensor features ranking according to SU measure values

Features	SU values	Features	SU values
Case temperature measurements		**Relative humidity measurements**	
V_CAS	0.216	AVR_RH	0.105
AVR_CAS	0.203	AF_DIFF_RH	0.086
AF_DIFF_CAS	0.141	FDIFF_RH	0.086
FDIFF_CAS	0.135	AS_DIFF_RH	0.079
AS_DIFF_CAS	0.103	SDIFF_RH	0.074
SDIFF_CAS	0.093	V_RH	0.053
CO_2 measurements		**PIR measurements**	
FDIFF_CO_2	0.280	TOS_PIR	0.327
AF_DIFF_CO_2	0.275	VTHI_LO_PIR	0.323
AVR_CO2	0.273	VTOS_PIR	0.293
AS_DIFF_CO_2	0.216	VTHI_LO_PIR	0.289
SDIFF_CO_2	0.174	AF_DIFF_TONP_PIR	0.160
V_CO_2	0.095	FDIFF_THI_LO_PIR	0.155
Ambient temperature measurements		**Sound measurements**	
AVR_TEMP	0.148	TOS_SND	0.385
AF_DIFF_TEMP	0.143	THI_LO_SND	0.370
FDIFF_TEMP	0.138	VTONP_SND	0.367
SDIFF_TEMP	0.065	VTLO_HI_SND	0.312
AS_DIFF_TEMP	0.060	VLO_HI_SND	0.305
V_TEMP	0.053	AF_DIFF_TOS_SND	0.187

Table 4: Dominant sound and PIR sensors features (in colours) from various features combination

n=1	n=2	n=3	n=4	n=5	n=6
V_CAS	FDIFF_CO_2	TOS_SND	TOS_SND	TOS_SND	V_CAS
FDIFF_CO_2	TOS_SND	TOS_PIR	TOS_PIR	TOS_PIR	TOS_SND
TOS_SND	TOS_PIR	THI_LO_SND	THI_LO_SND	THI_LO_SND	TOS_PIR
TOS_PIR	THI_LO_SND	THI_LO_PIR	AVR_CO_2	V_CAS	THI_LO_SND
	THI_LO_PIR	VTOS_PIR	VTOS_PIR		THI_LO_PIR
		VTONP_SND	VTONP_SND		AVR_CO_2
		AVR_CO_2	AS_DIFF_CO_2		VTONP_SND
			VTLO_HI_PIR		VTLO_HI_SND
			VTLO_HI_SND		VHI_LO_SND

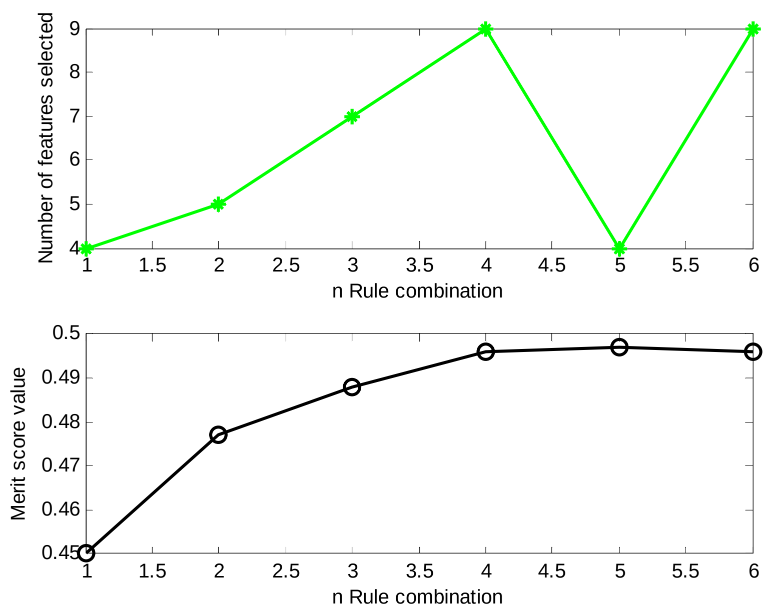

Figure 4. Merit score and different n Rule feature combination

A 10-fold cross validation was used for each n Rule combination, such that a dominant feature was selected once in each of the 10 runs. From figure 4, it is clear the number of features combination increased with merit score, indicating increasing predictive ability of selected feature subsets. For n=1, 4 features were selected with merit score of 0.450, increasing to 0.477 with 5 features selected at n=2. This trend continued at n=3 and n=4, with merit score of 0.488 and 0.496 respectively. However, the merit score increased marginally after a combination of 24 features (when n=4), by 0.001, and decreased again after 36 features (n=6) were combined. This may be due to over-fitting of the data. Further analysis by combining more features was stopped.

From table 3, sound and PIR features had the highest SU values, and hence the highest predictive power than other features. This provides an indication for their dominance in the n Rule combination process, as shown in table 4. This is not surprising, as the other features with good occupancy predictive strength such as case temperature and CO_2 features (even if occupants make up a significant CO_2 source in buildings) may not necessarily track occupancy numbers in the space. The large volume of the open-plan space coupled with air infiltration from large windows and ventilation hatches may have affected CO_2 features ability to track well with occupancy numbers, hence impacted on their dominance in the feature selection process. Case temperature features dominance in the selection process may have been weakened by the fact that not all occupants in the space make use of desktop computers, which were instrumented for case temperature measurements. Some

make use of their laptops. However, case temperature monitoring is unable to detect their presence. Non-selection of ambient temperature and relative humidity features may suggest that these parameters were dominated by building heating and cooling, and not necessarily due to occupancy.
A detailed description of CFS algorithm can be seen in (Hall, 1999). TOS_SND, TOS_PIR and THI_LO_SND are considered the three most dominate features, as per the number of times they were selected in the n Rule combinatory process. Figure 5 illustrates an overview of the data processing architecture.

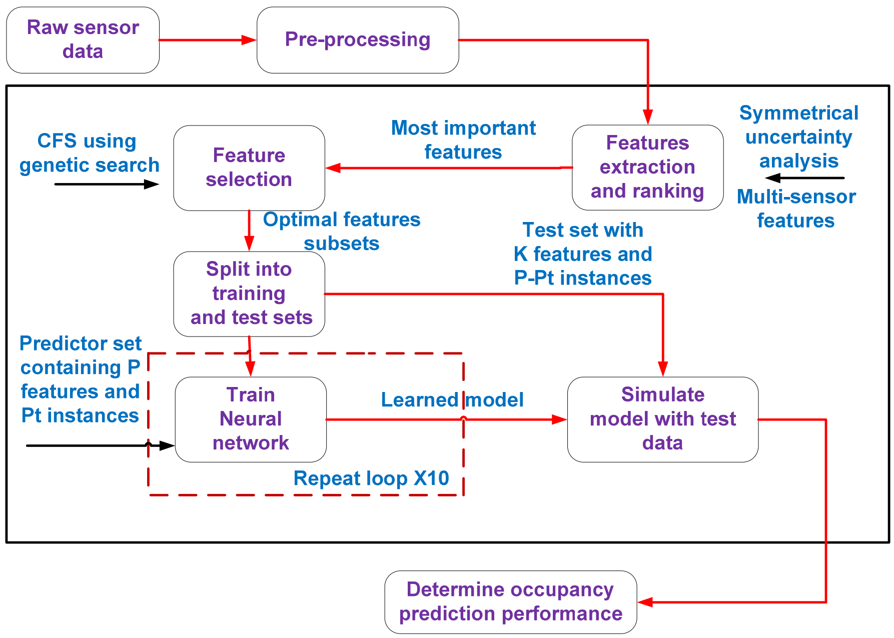

Figure 5. Overview of the data processing architecture

OCCUPANCY ESTIMATION RESULTS
In this study, a back propagation artificial neural network (ANN) was applied for occupancy numbers estimation. ANN based fusion model has been adopted for its fast computational capability and its relative simplicity to use. The model was implemented using the *MATLAB* Neural Network toolbox. Figure 6 shows the ANN architecture used for the study. TOS_SND, TOS_PIR and THI_LO_SND 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 was repeated for 10 times to increase the probability of reaching a global solution. The resulting average from the outputs of the training phases was used for analysis in the next stages.

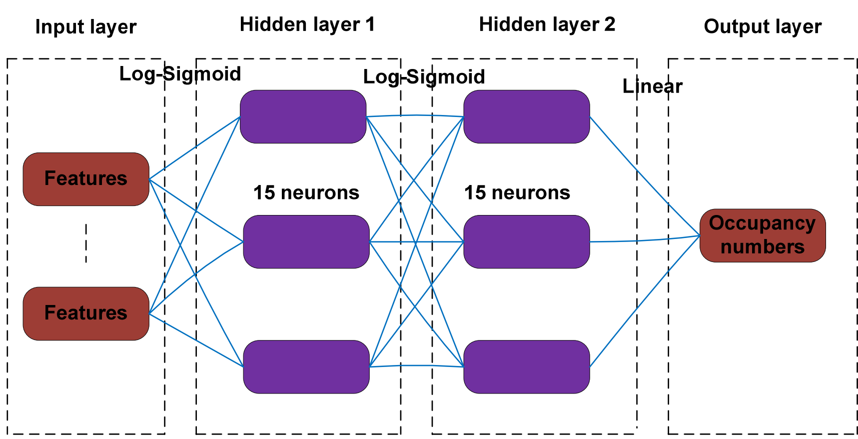

Figure 6. ANN Architecture used for the study

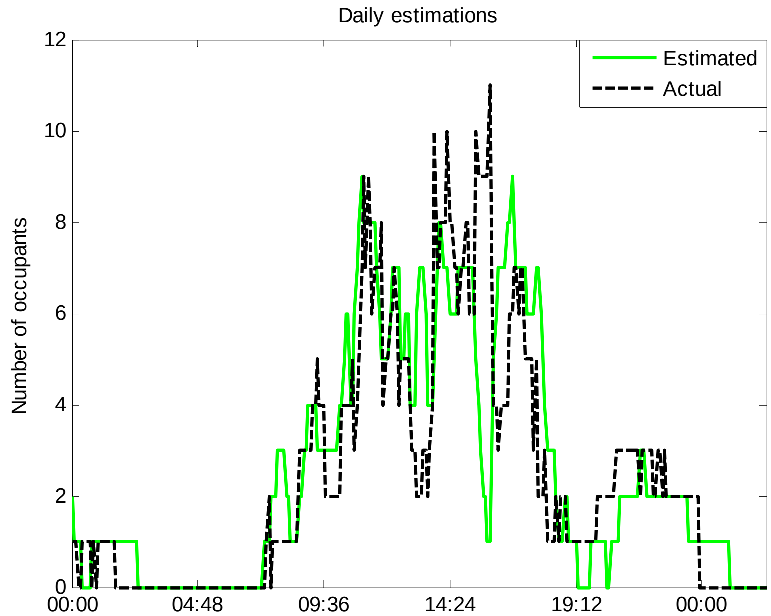

Figure 7. Occupancy estimation for 10/07/2012

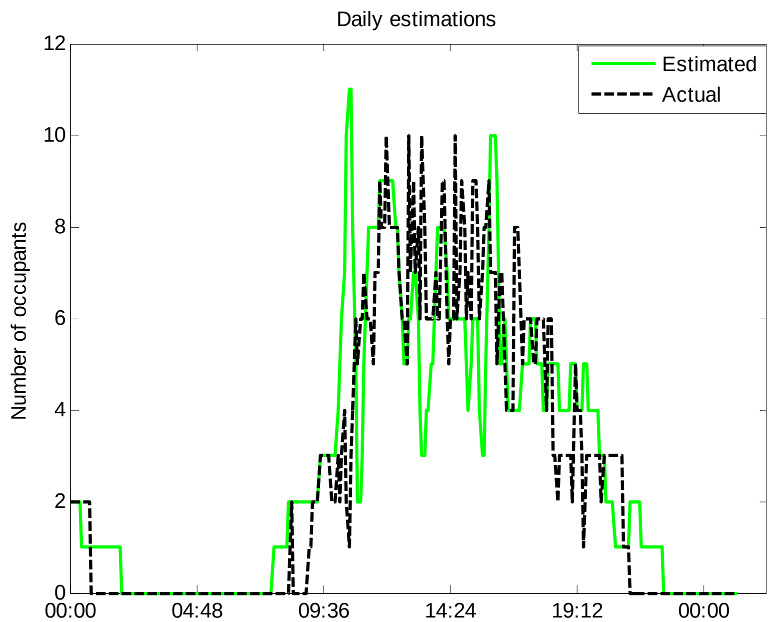

Figure 8. Occupancy estimation for 11/07/2012

The main statistics used for evaluating the model performance is the root mean square error (RMSE) and accuracy. RMSE measures the difference between estimated and actual occupancy data, while accuracy was obtained by dividing the number of correctly estimated data instances by the total number of data instances. For each daily result presented, the ANN model was trained using six days' worth of data, while the remaining day was used for model testing. As expected regular patterns in occupancy data are observable for the two typical days, as shown in figure 7 and 8. Occupancy levels were low in the early hours of the morning till about 8:00am, and steadily rise, peaking during lunch time between 12:00 and 4:00pm, where both staffs and PhD students (including occupants resuming late) would have arrived in most cases, if they are in for the day.

Occupancy levels decreased again, as occupants finish for the day and vacate the space. Some researchers normally stay long into the early hours of the morning. Hence, results show marginal occupancy levels during this period. Model estimations clearly track well with occupancy numbers, with an accuracy of 72.37% and RMSE of 1.58 on the 10/07/2012, and 70.45% and RMSE of 1.63 on the 11/07/2012. These results are considered good since the number of occupants in the test area varied between 0 and 11. This suggests that the model estimations were usually within 2 of the actual occupancy data; such difference may not cause any significant HVAC operation, unless the space switches from occupied to unoccupied, and vice versa.

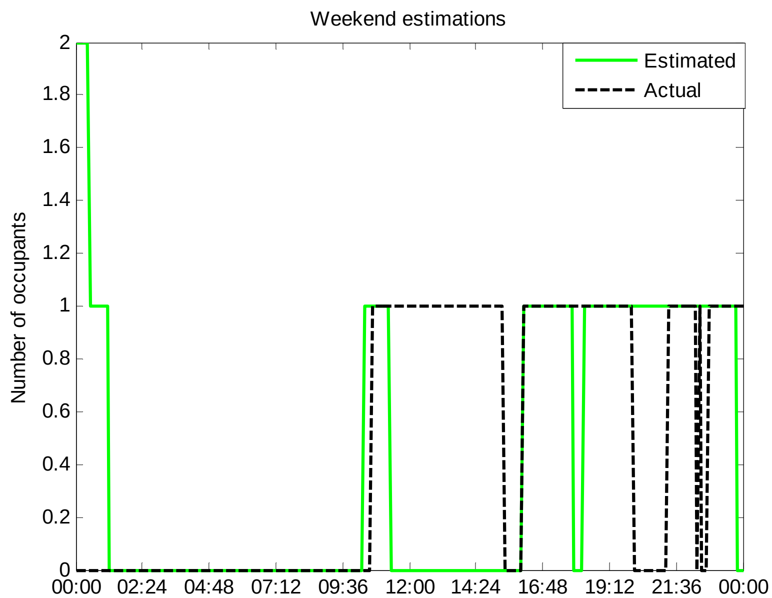

Figure 9. Occupancy estimation for 14/07/2012

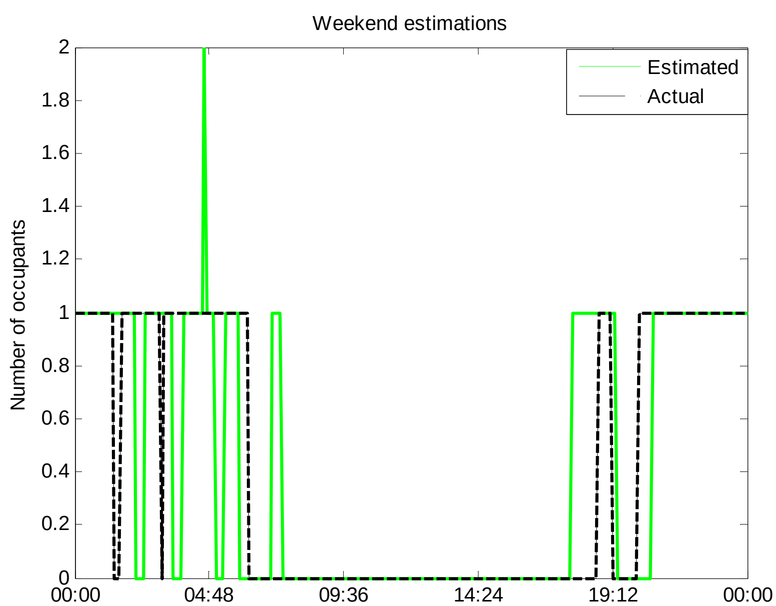

Figure 10. Occupancy estimation for 15/07/2012

For weekends, model estimations were also impressive, as shown in figure 9 and 10. Model estimations track actual occupancy levels with an accuracy of 68.11% and RMSE of 0.61on the 14/07/2012, and accuracy of

84.59% and RMSE of 0.37 on the 15/07/2012. These results may suggest the model efficacy for detecting space occupancy and vacancy, which may be crucial in any occupancy- driven HVAC operation.

Although, CO_2 sensors are widely used for regulating CO_2 levels especially for demand controlled ventilation (DCV) systems (Nielsen and Drivsholm, 2010, Kar and Varshney, 2009, Sun et al., 2011), this may not always be suitable for conditioning strategies, as CO_2 build-up is often slow, such that by the time sensors detect high levels of CO_2 that trigger ventilation, occupants may already be in state of discomfort (Fisk et al., 2006). Besides, CO_2 sensors can suffer significant drift over time (Shrestha and Maxwell, 2010), which may hamper the effectiveness of ventilation systems. Model results being obtained from information fusion using low-cost sound and motion sensors, may suggest the applicability of these sensors for occupancy numbers detection in a demand-driven HVAC control strategy in an open plan office, as they may serve as a far cheaper option than CO_2 sensors. However, further testing is recommended.

CONCLUSION

A novel methodology for estimation of occupancy numbers using symmetrical uncertainty analysis for feature selection has been presented. Results indicate that features from sound and motion sensors were the most effective for occupancy level estimation in the test area. Estimated occupancy numbers show reasonable tracking with actual occupancy levels. The model performance was impressive, with testing accuracy reaching 84.59%. Although, results are limited to the specific environment used in the study, it has the potential for wider applicability in other building environments, subject to further testing. Findings from the work may suggest the feasibility of exploring alternative sensors (such as sound and motion sensors) for occupancy number estimation in demand-driven HVAC control strategies. Future work could include exploring potential energy saving opportunities using occupancy estimates from the model in building control applications, and also testing this methodology with other standard machine learning approaches.

REFERENCES

ASHRAE. (2007). *ASHRAE 90.1 Standard*. Energy Standard for Buildings Except Low-Rise Residential buildings.American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc.

Benezeth, Y., Laurent, H., Emile, B., and Rosenberger, C. (2011). Towards a sensor for detecting human presence and characterizing activity. *Energy and Buildings 43*(2-3): 305-314.

Blum, A., and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence* 97: 245-271.

Brown, N., Bull, R., Faruk, F., and Ekwevugbe, T. (2012). Novel instrumentation for monitoring after-hours electricity consumption of electrical equipment, and some potential savings from a switch-off campaign. *Energy and Buildings* 47: 74-83.

Chenda, L. and Barooah, P. (2010). An integrated approach to occupancy modeling and estimation in commercial buildings. *American Control Conference (ACC),*Baltimore, MD, June 30- July 2.

Delaney, D. T., O'Hare, G.M., and Ruzzelli, A.G. (2009). Evaluation of energy-efficiency in lighting systems using sensor networks. *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy Efficiency in Buildings*, Berkeley,CA, USA, November 3.

Dodier, R. H., Henze, G.P., Tiller, D.K., and Guo, X. (2006). Building occupancy detection through sensor belief networks. *Energy and Buildings* 38(9): 1033-1043.

Dong, B., Andrews, B., Lam, K.P., Höynck, M., Zhang, R., Chiou, Y., and Benitez, D.(2010). An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network. *Energy and Buildings* 42(7): 1038-1046.

Erickson, V.L., Carreira-Perpinan, M.A., and Cerpa, A.E. (2011). OBSERVE: Occupancy-based system for efficient reduction of HVAC energy. *Information Processing in Sensor Networks (IPSN), 10th International Conference*, Chicago, IL, April 12-14.

Fisk, W. J., Faulkner, D., and Sullivan, D.P. (2006). Accuracy of CO2 sensors in commercial buildings: a pilot study. Technical report, LBNL. Available online from http://www.escholarship.org/uc/item/78t0t90v#page-1 [Acessed 04/05/2013].

Hall, M. (1999). Correlation based feature selection for machine learning. Dept. of Computer Science. New-zealand, University of Waikato. Doctoral dissertation.

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

Jianfeng, C., Jianmin, Z., Kam, A.H., and Shue, L. (2005). An automatic acoustic bathroom monitoring system. *Circuits and Systems, IEEE International Symposium,* Kobe, Japan, May 23-26.

John, G. H., Kohavi, R., and Pfleger, K., (1994). Irrelevant feature and the subset selection problem. Machine Learning: *Proceedings of the Eleventh International Conference*, New Brunswick, N.J., Rutgers University.

Kar, S. and P. K. Varshney (2009). Accurate estimation of indoor occupancy using gas sensors. *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 5th International Conference , Melbourne, December 7-10.

Kaushik, A. R. and Celler, B.G. (2007). Characterization of PIR detector for monitoring occupancy patterns and functional health Status of elderly people living alone at home. *Technology and Health Care* 15(4): 273-288.

Lam, K. P., Hoynck, M., Dong, B., Andrews, B.; Chiou, Y., Zhang, R. (2009). Occupancy detection through an extensive environmental sensork network in an open-plan office building. *Proceedings of the Eleventh International IBPSA conference*. Glasgow, Scotland, July 27-30.

Lam, K. P., Höynck, M., Zhang, R., Andrews, B., Chiou, Y., Dong, B., Benitez, D. (2009). Information-Theoretic Environmental Features Selection for Occupancy Dtection in Open Offices. Eleventh International IBPSA Conference. Glasgow, Scotland, July 27-30.

Li, N., Calis, G., and Becerik-Gerber, B. (2012). Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations. *Automation in Construction* 24: 89-99.

Martani, C., Lee, D., Robinson, P., Britter, R., and Ratti, C. (2012). ENERNET: Studying the dynamic relationship between building occupancy and energy consumption. *Energy and Buildings* 47: 584-591.

Meyn, S., Surana, A., Lin, Y., Oggianu, S.M., Narayanan, S., and Frewen, T.A. (2009). A sensor-utility-network method for estimation of occupancy in buildings. *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference* , Shanghai, December 15-18.

Nielsen, T. R. and C. Drivsholm . (2010). Energy efficient demand controlled ventilation in single family houses. *Energy and Buildings* 42(11): 1995-1998.

Padmanabh, A., Sougata Sen, M., Vuppala, S., Malikarjuna, V., Kumar, S., and Paul, S. (2009). A wireless sensor network based conference room management system. *In BuildSys '09: Proceedings of the First ACMWorkshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. Berkeley, CA, USA, November 4-6.

Pérez-Lombard, L., Ortiz, J., and Pout, C. (2008). A review on buildings energy consumption information. *Energy and Buildings* 40(3): 394-398.

Senthamarai Kannan, S., and Ramaraj, N. (2010). A novel hybrid feature selection via symmetrical uncertainty ranking based local memetic search algorithm." *Knowledge-Based Systems* 23: 580-585.

Shrestha, S., and Maxwell, G. (2010). Product Testing Report Supplement :Wall mounted carbon dioxide (CO2) transmitters. *National Buildings Controls Information Program*. Available online from http://www.iowaenergycenter.org/grant-and-research-library/product-testing-report-supplement-wall-mounted-carbon-dioxide-co2-transmitters/ [Accessed 04/05/2013].

Shuai Li, N. L., Burcin Becerik-Gerber, and Gulben Calis (2011). RFID-Based Occupancy Detection Solution for Opitimizing HVAC Energy Consumption . *The 28th International symposium on Automation and Robotics in Construction and Mining*, Seoul, Korea, June 29-July 2.

Sun, Z., Wang, S., and Ma, Z. (2011). In-situ implementation and validation of a CO2-based adaptive demand-controlled ventilation strategy in a multi-zone office building. *Building and Environment* 46(1): 124-133.

Tachwali, Y., Refai, H., and Fagan, J.E. (2007). Minimizing HVAC energy consumption using a wireless sensor network. *Industrial Electronics Society, 2007. IECON 2007. 33rd Annual Conference of the IEEE,*Taipei, Taiwan, November 5-8.

Tomastik, R., Lin, Y., and Banaszuk, A. (2008). Video-based estimation of building occupancy during emergency egress. *American Control Conference*,Seattle, WA, June 11-13.

Yu, L., Liu, H., and Guyon, I., (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research* 5: 1205-1224.