Title

User Presence Estimation in Multi-Occupancy Rooms Using Plug-Load Meters and PIR Sensors

Author
Shetty, Sindhu

Publication Date
2017-12-01

User presence estimation in multi-occupancy rooms using plug-load meters and PIR sensors

Sindhu S. Shetty *Student Member IEEE*, Hoang Duc Chinh *Member IEEE*,
Manish Gupta and S. K. Panda *Senior Member IEEE*
Department of Electrical & Computer Engineering, National University of Singapore, Singapore-117576
email: shettysindhu63@u.nus.edu, elehdc@nus.edu.sg, eleskp@nus.edu.sg

***Abstract*—In the built environment, the paradigm is shifting from providing uniform environment for all occupants to accommodating individual preferences through decentralised comfort devices enabled by IoT (Internet of Things) and ubiquitous computing. The effective implementation of localised and personalised comfort-enhancing and energy-saving strategies depends critically on the ability to detect the occupant presence in the immediate local environment. In this paper, estimating individual user presence in a shared office-space room using plug-load meters monitoring the power consumption of users' desktop computers and PIR (passive infrared) sensors at every user's desk is investigated. By extracting informative features from the data obtained and simple k-means clustering analysis, best-case presence accuracies of 89-99% and absence accuracies of 87-96% are achieved as validated by comparing with the ground-truth data for 4 different users.**

***Index Terms*—occupancy monitoring; smart buildings; data driven buildings; clustering; building energy efficiency; presence detection**

I. INTRODUCTION

Buildings consume around 30% of the total electricity consumption by end-use in Singapore and about 75% of this energy is consumed by lighting and air-conditioning systems [1]. The conventional air-conditioning systems cool the entire space irrespective of occupancy state and thermal preferences of the users in the space resulting in overcooling, thermal dissatisfaction and energy wastage. This energy consumption can be reduced while taking care of the comfort of the users by increasing the set-point temperature and providing cooling through personal fans or ventilation valves according to user's preferences [2]. Similarly, the energy consumed by lighting systems can be reduced lowering the intensity of the ambient lighting system and compensate this by providing personal task lighting only when required. The effectiveness of such smart systems and strategies greatly depends on the accurate detection of the individual user presence in his/her workspace and not just the occupancy estimation of the whole space.

Advanced communication technologies have enabled Wireless Sensor Networks (WSNs) to be deployed widely. A number of sensor node platforms have been developed and used in building environment such as TelosB, Micaz, Iris from Memsics, Arduino and Xbee, Waspmote from Libelium, etc. In this paper, the plug-load power consumption profiles and data from PIR sensors collected via WSNs are used to estimate user presence through a clustering analysis. Related work on occupancy monitoring in buildings and the motivation for this project is discussed in Section II. The data collection setup, sensors and meters used and the dataset obtained is described in Section III. The power profile features and count data from the PIR sensors are explored in Section IV and informative features are extracted. The k-means clustering analysis is detailed in Section V and the results are compared with the ground-truth. In Section VI, the effect of user behaviour on the accuracy, limitations of the current approach and methods to improve the estimation are discussed. Finally, the paper is concluded in Section VII.

II. RELATED WORK AND MOTIVATION

Most of the works in literature focus on total occupancy estimation of conference rooms, residential buildings or presence estimation in a single-occupancy office using supervised statistical methods. In [3], the authors use temperature, humidity, light and CO_2 measurements for occupancy detection using statistical classification methods. A static camera and video analysis is used for occupancy detection and people counting in [4]. The authors in [5] use digital electricity meters with ground-truth information from 5 households for occupancy detection. In [6], the authors propose *PresenceSense* algorithm for online presence detection based on plug-load monitoring. The authors in [7] propose occupancy detection methods using PIR, CO_2, door contact sensors and a camera in a shared office-space. The paper [8] gives a comprehensive review of existing works on occupancy monitoring for smart commercial buildings using smartphones, sensors, cameras, and RFIDs.

In this paper, plug-load monitoring and count data from PIR sensors is used to estimate the presence of individual users in a shared office scenario. Unsupervised learning method i.e clustering is used, hence the effort in obtaining the training labels is mitigated. New and existing buildings are being instrumented heavily with environmental sensors for assessing the indoor environment and meters to measure the energy consumption with the dropping cost and easy availability of IoT devices for this purpose. With office equipments and plug-loads consuming up to 33% of the total energy consumption in office and commercial buildings [9], it is reasonable to assume that the plug-loads will be metered in order to monitor their energy consumption and guide energy conservation measures. If the same meter data can be used as a proxy for occupancy estimation, this will reduce the need for having a dedicated occupancy estimation setup. In addition, inexpensive and eas-

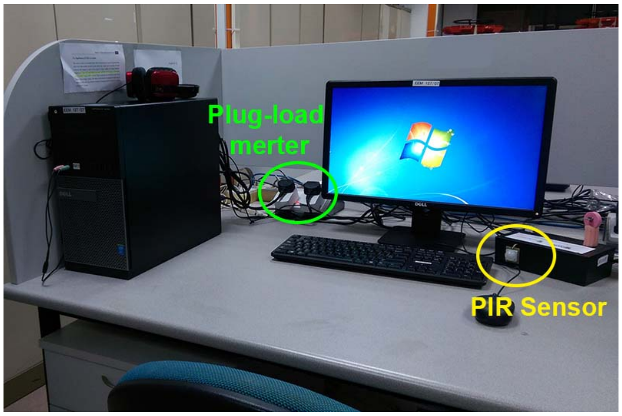

Fig. 1: Plug-load meters and PIR sensor deployment on a cubicle desk

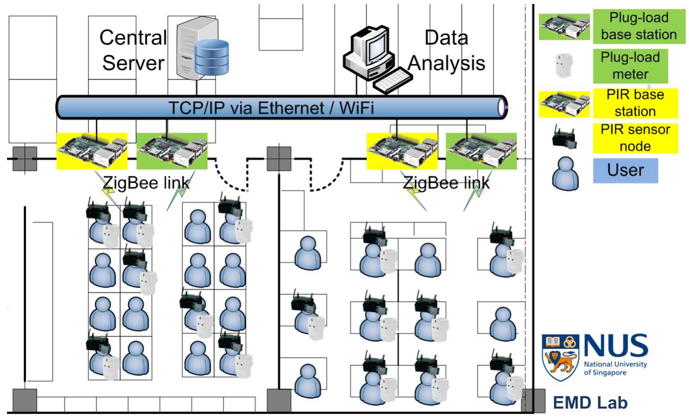

Fig. 2: Network architecture for data collection from plug-load meters and PIR sensor nodes

ily available PIR sensors which act as motion detectors are used to provide additional information. In brief, the proposed presence estimation method in this paper has the following advantages :

- Using unsupervised learning method is more practical as it can be very difficult to obtain the ground truth data in operational buildings.
- Using plug-load meters that are deployed for energy monitoring as proxy sensor for occupancy and inexpensive PIR sensors for additional information.
- Non-intrusive technique as opposed to using image or video analytics with camera which may have privacy concerns.
- Simple k-means clustering used as opposed to more sophisticated methods. This makes the analysis more interpretable and also allows for gaining additional insights through visualisation.

III. Data collection setup

In this work, a ZigBee based energy meter called Circle provided by Plugwise [10] has been adopted to monitor plug-load power consumption of electrical appliances in an office. The meter has the capability of acquiring the instantaneous power consumption and computing hourly energy consumption of the load. It is equipped with a ZigBee module for transferring data towards a base station as well as a relay to switch on/off the load. The meters are deployed on each cubicle desk to measure two appliances, i.e. the user's desktop personal computer (PC) and monitor as illustrated in Fig. 1. Besides, Fig. 1 also shows a PIR sensor placed on the same desk for motion detection of the user. This PIR sensor is connected to an Arduino board equipped with a Xbee Series 1 module.

All the meters have been deployed in two office rooms as shown in Fig. 2 and organized in two ZigBee networks, one in each room, in order to maintain good network connection. Each network consists of several meters and one base station. The base station is a single board computer, i.e. Raspberry Pi, equipped with a Plugwise ZigBee dongle called Stick to communicate with the Plugwise meters and is able to transfer data to a central server via Ethernet connection. The base station queries each meter in its ZigBee network in sequence periodically and reports to the central server if the measured value is changed. Similar to the meter network, the PIR sensor nodes in each are also grouped into a ZigBee network, detect the number of movement events within a fixed interval and sends the information to a base station, i.e. a Raspberry Pi board equipped with a Xbee module. The base station in turn forwards the information to the same central server.

As shown in Fig. 2, the overall architecture of the whole networks consists of four different ZigBee based networks used for connecting the base stations and sensors nodes meanwhile communication between the base station and the central server is implemented with TCP/IP via Ethernet. A MySQL database is deployed at the central server for storing collected sensor data together with a Web server application which provides RESTful services to retrieve data remotely for subsequent process and analysis.

Study cases of the presence estimation for 4 users in an operational shared office-space room are presented in this paper with our data collection framework. The power consumption of CPU and monitor of the user's workstation is monitored separately by two plug-load meters and the data is transmitted to the database every 30 seconds. A PIR sensor deployed on each user's workspace and the number of times motion detected in a 2 minute window is summed up and this count value is sent to the database. The ground truth data indicating if the user is actually present or absent at the workspace is recorded by observation for a period of one week.

IV. Exploratory data analysis

In this section, the count data from PIR sensors (PIR_{count}), CPU (P_{cpu}) and monitor (P_{mon}) power consumption data are analysed for each user along with the ground truth data. It is to be noted that the ground truth data in this case is utilised only for exploratory analysis (using ggplot2 [11]) and evaluating the performance. In practical scenarios, it is difficult and intrusive to obtain ground truth data for every user and

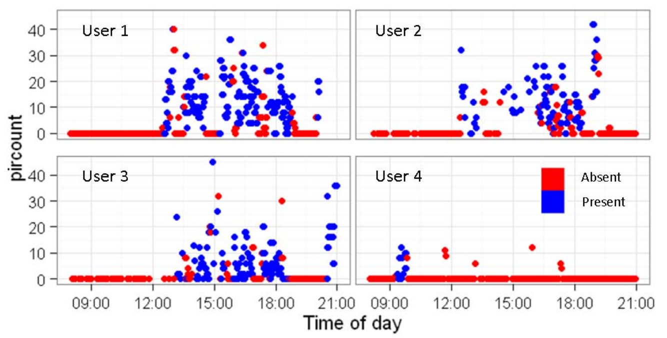

Fig. 3: Visualising PIR count data for the present and absent states of 4 users

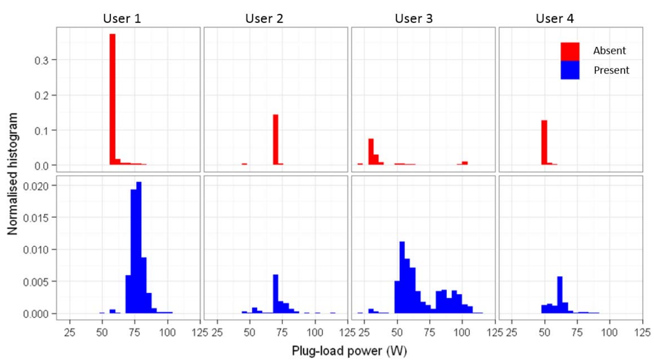

Fig. 4: Normalised histogram of plugload power for present and absent states for 4 users

hence unsupervised methods which do not require training data and labels are preferred over supervised methods.

Figure 3 shows the count data from PIR sensors placed at desks of 4 users along with the ground-truth for a single day. From the Figure 3, it can be noted that although the count data is higher for presence state than absence state, there are many present state points with value zero and absent state points with higher value. Hence, simple thresholding of PIR count data is not adequate for presence detection. Figure 4 shows the normalised histogram of plugload power consumption data ($P_{total} = P_{cpu} + P_{mon}$) for the users divided by the present and absent states. The power consumption levels and consumption patterns are different for different users and hence, a simple solution like using threshold levels to see if the user is present

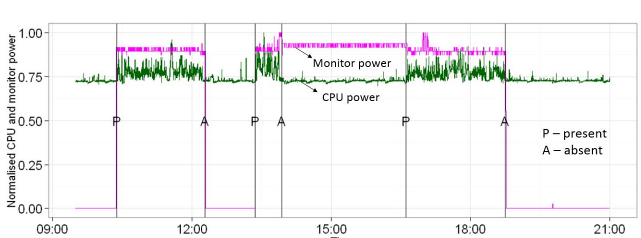

Fig. 5: Time series plot of normalised CPU and monitor power consumption for user 1 with annotated ground truth

User	State	P_{mon}(W)		P_{cpu}(W)		PIR_{count}	
		mean	sd	mean	sd	mean	sd
1	Absent	0.42	2.09	59.37	1.85	1.12	4.80
	Present	10.57	1.16	66.46	4.67	13.10	7.97
2	Absent	22.82	4.81	46.63	1.48	0.30	2.24
	Present	20.15	8.11	50.82	6.10	12.13	9.03
3	Absent	2.42	6.31	37.78	21.27	0.25	1.81
	Present	18.44	3.26	50.13	16.85	7.98	8.07
4	Absent	13.63	0.15	36.85	3.43	0.08	0.87
	Present	14.40	0.33	46.76	6.56	3.14	3.71

TABLE I: Summary statistics for the 4 users' data divided by present and absent states

will not work. The time-series plot of the CPU and monitor power consumption for user 1 on a typical working day along with the ground truth data is shown in Figure 5. From this plot, two types of behaviour can be observed - (a) the monitor is off when the user is not around and (b) the monitor is on when the user is absent. In both cases, the variance/standard deviation of the CPU is higher during present period than the absent period. Different users are likely to exhibit different behaviours - some may keep the monitor on when not around, some shut down the computer during non-working hours while others put it on sleep mode. It is necessary to account for the user behaviour or the pattern of plug-load usage. The mean and standard deviation for the CPU power, monitor power and count data from PIR sensors is shown in Table I so that these values can be compared for the present and absent states. From this Table I, it is clear that users 1 and 3 have a similar behaviour (although with different values of data) whereas the monitor power does not give much information about the presence for users 2 and 4. The standard deviation values of cpu power consumption and count data from PIR sensors are the most consistent features across the 4 users with lower values for absent state and higher values for present state, again with better distinction for some users than others. It is also worth noting as the PIR sensors are placed on each user's desk, the relevance of this data depends on its position and location of the user's workspace and hence, it may provide less information about presence as observed in the case of user 4 when compared to other users.

From the exploratory analysis, it is clear that unsupervised presence estimation method will perform differently based on which features are used for the analysis and will depend on the user behaviour. The learning model or analysis must be personalised to account for this and a single rule will not work for all users. The following features are extracted from the data for each user and the performance is evaluated for different feature combinations as described in Section V. The features are extracted for PIR_{count}, P_{total}, P_{cpu} and P_{mon} each of which is represented by X_i. The rolling mean and standard deviation are calculated with a fixed window length (in this case 10 samples which accounts for a 5 minute interval) and window alignment at the center i.e it considers both past and future values as this is an offline analysis.

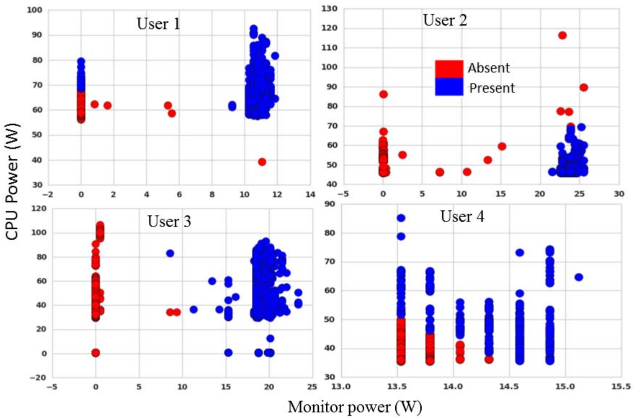

Fig. 6: Visualising the clusters for feature combination of monitor power and CPU power

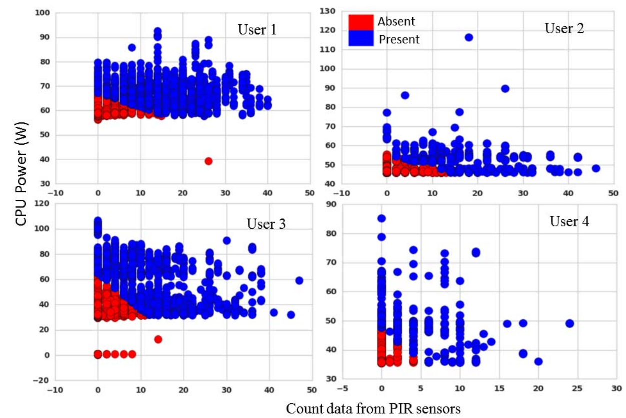

Fig. 7: Visualising the clusters for feature combination of count data from PIR sensor and CPU power

- Rolling mean of m^{th} value $X_{i,m}$ with fixed window w

$$\bar{X}_{i,m} = \frac{1}{w} \sum_{k=m-\frac{w}{2}}^{m+\frac{w}{2}} X_{i,k}$$

- Rolling standard deviation of m^{th} value $X_{i,m}$ with fixed window w

$$SD_{i,m} = \sqrt{\frac{1}{w-1} \sum_{k=m-\frac{w}{2}}^{m+\frac{w}{2}} (X_{i,k} - \bar{X}_{i,k})^2}$$

V. Clustering analysis

Once we have obtained the data, the instances can be clustered into two groups (which is interpreted as the absent and present state by visualising the clusters). To solve the clustering problem, one of the simplest and popular method namely k-means clustering is used [12]. In this analysis, we use the k-means algorithm implementation in the Python package *scikit-learn* [13]. The k-means method is applied to different feature combinations for the 4 users' data. Figures 6 and 7 show few representative clustering results out of the many cases analysed. From the exploratory analysis, it is known that the feature values typically have lower values for the absent state than the present state. Hence, by visualising the clusters, the cluster with mostly lower values of features is assigned to the absent state and the other cluster to the present state. The assigned clusters are then compared with the actual ground truth data and the results are shown in the Table II. Since the percentage of present and absent cases in the collected data are different, it is crucial to look at the present and absent accuracies separately rather than the overall accuracy.

Features	User 1			User 2		
	O	Pr	Ab	O	Pr	Ab
M, C	96.1	98.9	95.7	10.2	84.9	4.3
Cm, Cs	93.3	91.8	93.5	92.3	46.3	95.9
Tm, Ts	94.6	98.5	94.1	9.7	70.7	4.9
C, P	93.2	85.7	94.2	95.4	55.6	98.5
Cs, P	92.2	90.4	92.4	94.0	51.6	97.3
P, Ps	91.8	71.2	94.8	95.6	78.1	97.0
Pm, Ps	92.2	75.4	94.6	95.9	89.1	96.5
M, C, P	96.5	98.0	96.3	91.9	54.9	94.8
Features	User 3			User 4		
	O	Pr	Ab	O	Pr	Ab
M, C	90.3	97.1	87.7	97.0	88.2	97.8
Cm, Cs	75.4	53.2	83.8	93.3	93.2	93.3
Tm, Ts	78.6	68.4	82.5	93.4	93.6	93.4
C, P	80.6	55.2	90.2	95.6	74.8	97.5
Cs, P	83.0	55.8	93.2	90.0	47.2	93.7
P, Ps	85.0	53.0	97.2	93.5	50.5	97.3
Pm, Ps	87.1	61.4	96.8	93.3	51.2	97.0
M, C, P	90.1	97.2	87.5	96.9	89.6	97.6

Legend					
O	overall accuracy (%)	Pr	presence accuracy (%)	Ab	absence accuracy(%)
M	P_{mon} (W)	C	P_{cpu} (W)	P	PIR_{count}
Cm	$\bar{P}_{cpu}$ (W)	Cs	$SD(P_{cpu})$ (W)	Tm	$\bar{P}_{total}$ (W)
Ts	$SD(P_{total})$ (W)	Pm	$\bar{PIR}_{count}$	Ps	$SD(PIR_{count})$

TABLE II: Comparison with ground truth for various feature combinations

Looking at Table II, for user 1, the accuracies are consistently over 90% for all feature combinations except a lower presence accuracy when using only the PIR_{count} data. For user 4, the performance is similar to user 1 but with a significantly lower presence accuracies while considering combinations with PIR_{count} data. For user 3, using P_{mon} and P_{cpu} data gives the best results with 97% presence and 88% absence accuracy. In case of user 2, using the PIR_{count} only, gives the best accuracy at 96% overall with 89% present and 96.5% absent accuracies. Note that using the P_{mon} data for user 2, it is not possible to get the absent states right as informed by the exploratory data analysis earlier.

VI. Discussion

In this section the impact of user behaviour on the presence estimation methods and ways of increasing the accuracy are discussed.

A. Note on user behaviour

From the performance Table II, it is clear that the accuracy of the proposed method depends greatly on the user behaviour and the placement of PIR sensors that can provide relevant additional information. Clues about user behaviour and placement of sensors can be obtained by another clustering analysis prior to the final clustering to obtain the present and absent states. In this case, the elbow method is used to find the optimal number of clusters by minimising the WCSS (within-cluster sum of squares) metric for the two feature combinations namely, monitor power along with CPU power data as shown in Figure 8 and count data from PIR along with CPU power data as shown in 9. This strategy assumes that the user behaviour remains the same throughout the data collection period. Using the intuitive elbow method, the suggested number of clusters is k if there is a sharp decrease of WCSS for k clusters and no significant decrease for $k+1, k+2, k+3...$ clusters. In our case, if the WCSS plots in Figure 8 for a user shows the optimal clusters is 2, then features from monitor power and cpu power data can be used to give good results. The plots for user 1 and user 4 exhibit this behaviour and this is validated by the good accuracies for user 1 and user 4 in Table II. Looking at the WCSS plots in Figure 9, if the optimal number of clusters is 2, then the count data from PIR sensors will be a significant informative attribute. This is confirmed by the results for user 2 in Table II which greatly improves for features involving PIR_{count} data only. In the case of user 3, the suggested number of clusters by elbow method is 4 for feature set of P_{mon}, P_{cpu} and 3 for feature set of PIR_{count}, P_{cpu}. The clustering is carried out with the suggested number of clusters and the results can be seen in Figure 10. In both sets of feature combinations, the group in red (Cluster 1) contains higher range of P_{cpu} values and lower range of P_{mon} and PIR_{count} values. This behaviour is odd and could be interpreted as the user using the CPU on remote access when in absence state. However, this does not affect the performance of the presence estimation method as evident in Table II for user 3 accuracy results with P_{mon} and P_{cpu} features. This is because the remote access cluster is correctly identified as absence state for user 3 as seen in Figure 6. Thus, investigating the WCSS plots for different feature combinations provides insights on user behaviour and informs the best set of features for a particular user which can then be used in the final clustering process with 2 clusters.

B. Suggestions for improvement of the methods

There are several ways to improve the accuracy of the presence estimation method. Data can be collected more frequently, this may improve the accuracies marginally because of the reduction in delay due to the current sampling interval of 30 seconds and 2 minutes for plug-load power and PIR sensors respectively. If the ground truth data is available, the size of the window for rolling statistics can be tuned to obtain the best separation between the present and absent states. As the performance depends on user behaviour, effort in obtaining the behaviour information directly from user through surveys may pay off in informing the best set of features. Finally, time of day information is not included in the analysis. As we can expect each user to follow some routine even if the working hours are flexible or the working hours are fixed for all users in the office, in both cases, including time of day related data could improve the performance of the presence estimation methods.

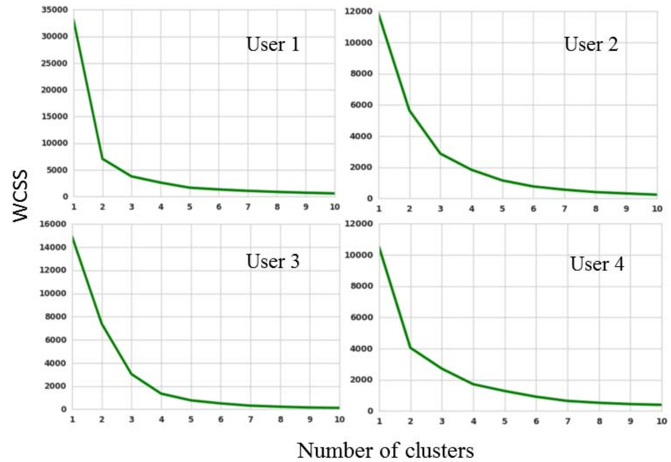

Fig. 8: Plot of WCSS against the number of clusters for the features P_{mon} and P_{cpu}

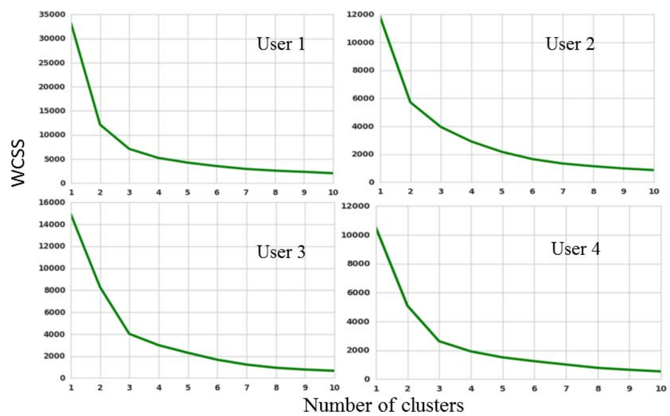

Fig. 9: Plot of WCSS against the number of clusters for the features P_{cpu} and PIR_{count}

VII. Conclusion

In this paper, the plug-load power consumption data and the count data from PIR sensors are used to estimate the individual presence of the user in a shared office-space. Clustering analysis is used for both inferring user behaviour and then assigning the data points into present and absent states based on the most relevant combination of features. The proposed method is compared with the ground truth recorded for 4 users and it achieved 89-99% presence and 87-96% absence accuracies for the best set of features. The future work is focused on online presence detection which can then be used in many real-time control strategies in which the

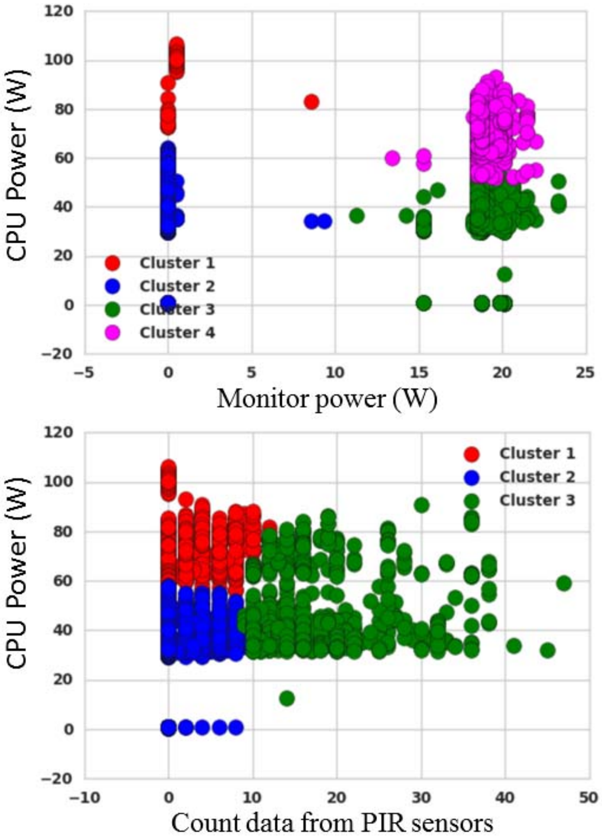

Fig. 10: Clustering results for user 3 with suggested number of clusters by the elbow method : (a) with features P_{mon} and P_{cpu}, (b) with features PIR_{count} and P_{cpu}

insights obtained in the above analysis can guide the methods. Another interesting application would be energy consumption feedback systems that attempt to induce behavioural change in the users towards energy-savings in which case the presence detection methods will also have to adapt to change in the user behaviour. Overall, it is demonstrated how the data collected from IoT devices along with data analytics can be used to gain insights on the user presence and energy usage behaviour in operational office-space buildings.

ACKNOWLEDGEMENT

This research is funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore. The authors would also like to thank the Electrical Machines and Drives Laboratory(EMDL) group of National University of Singapore (NUS) for taking out time to participate in the experiment and we value their patience during the experiment period.

REFERENCES

[1] K. Chua, S. Chou, W. Yang, and J. Yan, "Achieving better energy-efficient air conditioning–a review of technologies and strategies," *Applied Energy*, vol. 104, pp. 87–104, 2013.

[2] S. Schiavon and A. K. Melikov, "Energy saving and improved comfort by increased air movement," *Energy and buildings*, vol. 40, no. 10, pp. 1954–1960, 2008.

[3] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and co 2 measurements using statistical learning models," *Energy and Buildings*, vol. 112, pp. 28–39, 2016.

[4] Y. Benezeth, H. Laurent, B. Emile, and C. Rosenberger, "Towards a sensor for detecting human presence and characterizing activity," *Energy and Buildings*, vol. 43, no. 2, pp. 305–314, 2011.

[5] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, "Occupancy detection from electricity consumption data," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, pp. 1–8, ACM, 2013.

[6] M. Jin, R. Jia, Z. Kang, I. C. Konstantakopoulos, and C. J. Spanos, "Presencesense: Zero-training algorithm for individual presence detection based on power monitoring," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, pp. 1–10, ACM, 2014.

[7] H. B. Gunay, A. Fuller, W. O'Brien, and I. Beausoleil-Morrison, "Detecting occupants' presence in office spaces: a case study," in *eSim 2016, Hamilton, ON*, 2016.

[8] K. Akkaya, I. Guvenc, R. Aygun, N. Pala, and A. Kadri, "IoT-based occupancy monitoring techniques for energy-efficient smart buildings," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2015 IEEE*, pp. 58–63, IEEE, 2015.

[9] "Assessing and reducing plug and process loads in commercial office and retail buildings." http://www.nrel.gov/docs/fy13osti/54175.pdf. Accessed: 2017-03-29.

[10] "Plugwise energy meters." https://www.plugwise.com/products/appliances-and-lighting/energy-meters-and-switches. Accessed: 2017-03-27.

[11] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[12] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297, Oakland, CA, USA., 1967.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.