Taking advantage of an existing indoor climate monitorization for measuring occupancy

UNAI SARALEGUI
TECNALIA Research & Innovation
Sustainable Construction Division
Area Anardi 5, 20730 Azpeitia
SPAIN
usaralegui@gmail.com

MIGUEL ÁNGEL ANTÓN
TECNALIA Research & Innovation
Sustainable Construction Division
Area Anardi 5, 20730 Azpeitia
SPAIN
mangel.anton@tecnalia.com

JOAQUÍN ORDIERES-MERÉ
Universidad Politécnica de Madrid
ETSII, PMQ Research Group
José Gutierrez Abascal 2, 28006 Madrid
SPAIN
j.ordieres@upm.es

Abstract: This paper describes a procedure to gain additional information from an already existing infrastructure primarily designed for other purposes. The deployed sensor network consists of wirelessly communicated indoor climate monitoring sensors, for which it is tried to extend its usage by determining occupancy in the room they are located, in that way the system provides a higher level aspect of the house usage. An elderly caring institution's building has been monitored for one year obtaining data about temperature, relative humidity and CO_2 levels from five different rooms. Such data shows some interesting patterns as the air flow between the rooms which should be considered in any real case scenario. The data has been used to train some machine learning models, which show acceptable quality overall suggesting to use this kind of sensing equipment to perform an occupancy monitoring non-intrusively. The acquired knowledge could bring additional opportunities in the care of the elderly, especially for specific diseases that are usually accompanied by changes in patterns of behaviour. By using the occupancy status it could be possible to determine changes in the daily patterns in that segment of the population which could be an indicative of the initial states of a disease or a worsening in it.

Key–Words: Domestic occupancy, Smart Buildings, Climate sensors, Internet of Things, Pattern analysis, Health Monitoring, Machine Learning

1 Introduction

Recent advances in the Internet of Things (IoT) field have been a great impulse in the smart building and cities field, also for occupancy and pattern analysis areas. Those advances have helped to monitor the environment in multiple ways, but there are still some limitations because some variables, such as the occupancy, are not possible to measure directly, in that sense several researches have been carried out proposing to use various sensors to estimate the occupancy from the sensor's data.

Determining the occupancy in the rooms of a building is interesting for multiple purposes, as energy efficiency or pattern analysis. This paper focuses in determining the occupancy for finding patterns in the movements and the time table of the house host in order to detect possible symptoms of diseases associated with dementia. In that sense, the proposed approach is defined as an ambient assisted living (AAL) framework, which stands for systems that make use of information and communication technologies to develop applications and services for the elderly, in order to help them in their daily activities [1].

Elderly people who live alone usually suffer more accidents than other people. For example, in the district of Kaiserslautern, Germany, 30 per cent of people who are more than 65 years of age and live alone at home suffer at least one fall per year. Therefore, being able to identify either sudden or progressive changes in their routines by using an already existing infrastructure, or an easily installable and a non-intrusive one, could help in providing specific care to such a population segment.

As the aim is to design an easily installable and non-invasive system, an already existing monitoring of this type has been used to obtain the data from five rooms in an elderly caring institution's building. The sensors in each room measure temperature, relative humidity and CO_2 concentrations and are primarily designed for HVAC controlling.

As we seek to monitor the behaviour of the elderly to detect changes in their patterns of behaviour, it is important to be able to record the movements throughout the building in order to detect changes in the everyday patterns that may indicate dementia. With that objective in mind, we show in this paper how to detect human presence in a room by using the sensors that are already deployed to monitor the air quality.

In order to determine the occupancy, a data analysis has been carried out for inferring relations between the variables and the occupancy, while machine learning algorithms have been used in order to build binary classifiers which output a prediction for the occupancy of a room based on the monitored data.

For better understanding the work being done, the structure of the paper will first introduce the problem by analysing the present state of the art in the following section. The data acquisition process and the used modules are shown in the third section. The fourth section will provide the proposed methodology. Section five will introduce the data analysis performed over the acquired data. Next, on section six some machine learning models are trained for determining binary presence showing the results obtained. Finally, the last section will give the main conclusions of the work and discuss future developments and research.

2 State of the Art

The use of IoT technologies has made able to acquire greater knowledge of different behaviours enabling the spread of research paths in this ambit. In this area, studies of air quality and energy efficiency in buildings have received much attention as in [2, 3] where indoor monitoring is proposed and occupancy is related to comfort measurements.

Indoor air quality monitoring has been deeply investigated, where contributions like [4] propose the development of a compact battery-powered system that monitors the carbon dioxide level, temperature, relative humidity, absolute pressure and intensity of light in indoor spaces, and sends the measurements by means of the existing wireless infrastructure based on the IEEE 802.11 b/g standards. The idea of promoting low cost solutions instead of those that require extensive deployment was deeply investigated in [5, 6]. The latter used a monitoring system that was based on an Arduino platform with six sensors.

Applications of the knowledge about the building occupancy are various and diverse. For example, data about the occupancy is used for energy management of buildings, as the occupation data shows whether the rooms in an office are used or not and can also lead to build models that are able to predict the occupancy based on historic data in order to design energy efficient approaches [7, 8].

Monitoring the indoor climate status is not only valid for energy efficiency or comfort-related topics, but they can also be used for monitoring pollutant levels in indoor areas, which have shown to have bad effects in the users of the building. A correlation between the concentration of pollutants and health problems in schoolchildren has also been shown in some studies [9, 10]. Researchers in this area have also been able to link sleepiness or an impact on health to CO_2 levels on campuses or in offices, thereby showing the physiological effects that a high concentration of CO_2 indoors has on workers or students [11, 12].

Another trend more closely related to our research interest involves using an air quality monitoring system to derive occupancy information in a non-intrusive scenario. The key aspect is to work under long term service principle. Even if different approaches have been presented, they carry several limitations. One of the proposed approach is video-image based detection, as in [13], which is able to obtain good results but it is seriously limited as it carries strong privacy issues and computational requirements. Another existing approach is the one that makes use of passive infrared (PIR) sensors, as in [14], this method does not have any privacy related concerns but it is somehow limited as it requires movement to trigger. Those limitations have caused CO_2 based presence detection methods to receive increasing attention.

In our area of research, mixed approaches, such as [15], have been presented. They are based on analytical models that were calibrated on empirical data, with a decision tree that defines the final inference model for occupancy. Other authors have proposed the use of a Hidden Markov Model (HMM) as a convenient and effective approach for occupancy estimation by use of Multinomial Logistic Regression (HMM-MLR) [16]. Some research to determine binary occupancy information has also been undertaken, as in [17] where a binary presence detection framework is proposed using an indoor weather station's data and Hidden Markov Models. The approach in this paper will benefit from artificial intelligence based techniques to infer presence in a room using climate data recollected in a real scenario.

As it has already been mentioned, the occupancy estimation is aimed to be used for being able to detect changes in the everyday patterns in elderly people. The use of IoT devices for such a purpose is a widely extensive topic as it could aim to monitor the health status of the patients online [18]. Some research have already been done in this sense, as [19] which proposes a home monitoring system for patients with neurological disorders by monitoring body temperature, MRI images, hand movement trajectory, etc.

3 Data Acquisition

The main goal here is to be able to determine if a room is occupied or not by using indoor climate monitoring data, which usually consists of temperature, relative humidity, and in some cases CO_2 levels. With that aim in mind, some data need to be recollected in order to perform an analysis of the relation between those variables, and most importantly data is needed to build the models that will predict if the room is occupied or not.

As the objective is quite hard to achieve the collected data needs to be sufficient, and it must be taken in different months due to the climate changes through the seasons during a year. The data used for performing this research contains entries of an entire year, with one entry every ten minutes, and the acquisition has been performed respecting the national laws regarding non-intrusiveness and confidentiality.

An agreement with an elderly caring institution has allowed us to obtain data from an already deployed indoor climate monitoring sensor network, which was primarily designed for HVAC controlling. Such sensor network compliments our requirements of an easily installable and non intrusive system. The data of the monitoring of five different rooms has been used in this study, such rooms consisting in the main hall, the living room, the dining room, and two bedrooms (named room 1 and room 2). Figure 1 shows one of the sensor systems deployed in the building.

Five identical sensor systems were deployed in the rooms, one in each. The sensor systems consist in sensors which are capable of measuring temperature, relative humidity and CO_2 levels. Due to the energy requirements and with the aim of performing the monitoring without making changes in the infrastructure, the sensor systems are powered by ambient light by means of an energy harvesting solar cell. In prolonged periods of low light, security backup batteries provide power to the sensors. Cumplimenting the requirements the sensor systems do not use any type of cables for communications, and they send the information to a central node by using the open EnOcean® standard protocol (ISO/IEC 14543-3-10) for wireless communications [20].

As mentioned the dataset consists of a year of monitoring with one value taken every ten minutes. A central node is the responsible for obtaining such data and uploading it to a database in the cloud. A schematic version of the data acquisition modules is shown in Fig. 2. It has to been mentioned that thirteen residents were normally living in the building while the data was being collected, with the monitored rooms being empty in some cases. Due to the condition of the residents, their routine is well defined and maintained throughout the year, the benefits of this are that having the knowledge of the time table, provided by the institution, it is easier to know where the residents are located at a given hour, so abnormalities are easier to detect. In general, the bedrooms are used during the night and are empty during the day. The residents remain in the living room during the day, except at breakfast, lunch and dinner times when they congregate in the dining room.

Our goal is to develop a system which is easily installable or, even better, to use an already existing one, as it is done in this study. The system used was primarily designed for HVAC controlling that does not require high sampling frequencies, in this case the frequency is 10 minutes. It is clear that such sampling frequency limits the available information especially

Fig. 1: Sensor system located in one of the rooms (upright corner).

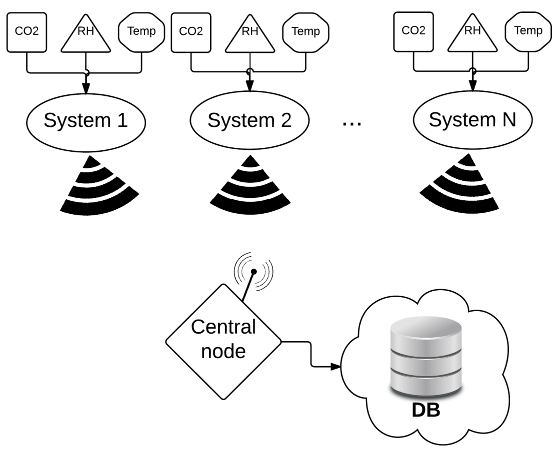

Fig. 2: Schematic of the data acquisition modules.

in great changes periods, that is why in an installation precisely dedicated for occupancy monitoring the sampling frequency should be higher. Such improvement in the data acquisition modules could lead to improvements in the accuracy of the developed system as the knowledge obtained would be higher and more controlled. Even if the monitoring frequency is slow it is shown that it is enough for determining binary occupancy.

4 Proposed Methodology

Even if this paper mainly focuses in the occupancy detection, the construction of the framework for monitoring the elderly for finding symptoms of dementia diseases is always behind. With this objective in mind, a framework for detecting those changes is proposed (see Fig. 3).

The proposed approach will make use of the explained data acquisition methodology, while additional sensors can be placed if they are considered to be relevant. In this case of study five rooms have been monitored, but in a real use case, in a domestic house for example, that will not be necessary, it will possibly be enough with fewer locations.

As a framework and in concordance with the Internet of Things paradigm the data will be uploaded to the cloud, in this way the data will be accessible from any place, which makes possible to work on the data, in this case to model the patterns of behaviour, remotely. That modelling will be performed by using artificial intelligence techniques, which will be helpful for extracting the desired knowledge. But not only will artificial intelligence techniques be used for the ultimate goal, but some previous steps, such as the occupancy detection explained in this paper, rely on machine learning algorithms to infer the occupancy from environmental data.

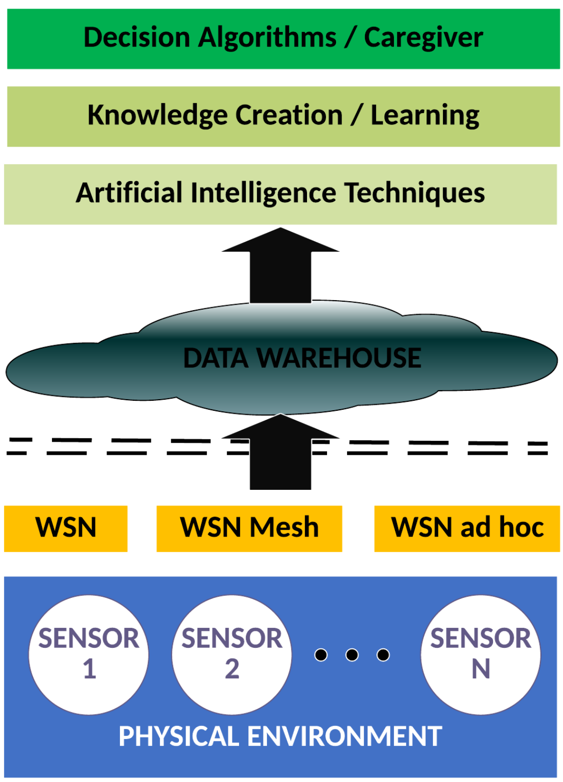

Fig. 3: Proposed methodology for the desired system.

This research is beneficial as it shows that it is possible to use an already available, or an easily installable, and non-intrusive sensing equipment to gain additional information. It will be also interesting to determine the best set of machine learning classification algorithms to predict the occupancy values. To this end, the problem is treated as a supervised learning process, in which linear and non-linear techniques will be used with the models, in a first approach, being built for each monitored room.

5 Data Analysis

As previously mentioned, an elderly caring institution has allowed to perform a monitoring in one of its buildings. The data acquisition has lasted for a whole year with one entry being saved in a database in the cloud every ten minutes. As commonly happens in real world problems some non-desired values, such as extremely high temperatures or extremely low humidity levels (not possible in the region where the building is located) also appeared in the dataset. Because of all these facts some data cleaning and pre-process was required to do. This could be a lack in the proposed system, but as the bad values are few in comparison with the size of the database it does not seem to be a critical problem.

A quick visualization of the data already enables to see some interesting relations. It is clear that the CO_2 levels are a valid indicative to determine the presence as it has been already reported in the literature [8, 17, 15, 21]. Figure 4 shows the evolution of the CO_2 levels in the monitored rooms for a given day, as it can be seen the CO_2 concentrations are higher in those places that are thought to be occupied, as the bedrooms at night or the dining room at breakfast, lunch and dinner times. Another notorious effect is the one that appears in the nights in the two monitored rooms, it can be seen how the concentration continually increases in room 2, while the concentration remains more stable at room 1 indicating that the door has been opened there, that way allowing the air to

Unai Saralegui, Miguel Angel Anton,
Joaquin Ordieres-Mere

flow to the adjoining spaces. This phenomenon has not been taken into account until now, but it seems to have a great importance as it could lead to false positives seeing the importance the CO_2 concentration has for determining presence.

The air flowing phenomena does not only appear when the room doors are opened, but it is a common phenomenon as it can be more clearly seen in Fig. 5. If such figure is analysed it can be seen that the CO_2 concentration in the rooms is affected by the concentrations in the bigger spaces, such as the dining room or the living room. An increase in the levels in the room 1 can be observed when the concentration increases in the dining room, while a similar phenomenon is observed between the room 2 and the living room. It has to be mentioned that such pairs of rooms are one in front of the other in the real scenario. These cases are not isolated ones, but the phenomenon repeats through the time, so it should somehow be taken into account, to avoid false positives derived from the air flow.

It is clear that the CO_2 concentrations will have a great importance for determining occupancy in a room, but on the contrary, the relative humidity and temperature values do not seem to give much information, in a visual analysis at least. Even if those values do not seem to give much information about the presence, the Fig. 7 shows that it is possible to detect when the windows are opened in the rooms. As it can be seen there is a huge peak around 9 a.m. which is the time when the institution's staff cleans and ventilates the rooms. As for the temperature evolution shown in Fig. 6, the values remain quite stable and do not show any changes that could be influenced by the presence of a person in a room, so they do not seem to be a valid indicative to measure the occupancy of a room.

In order to make the assumptions about the CO_2 concentration values stronger, the boxplots per month

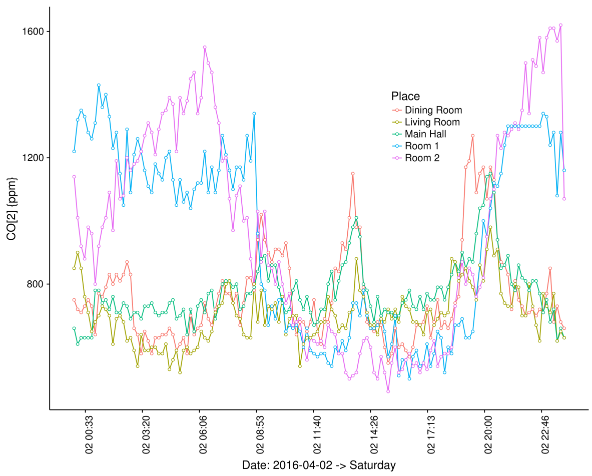

Fig. 4: CO_2 concentration levels (ppm), 2/4/2016.

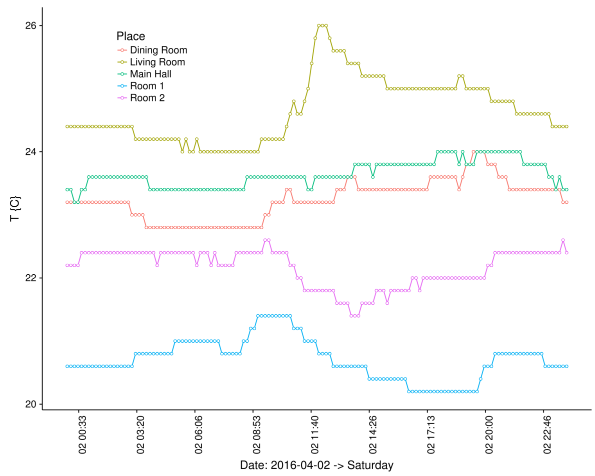

Fig. 6: Temperature values (°C), 2/4/2016.

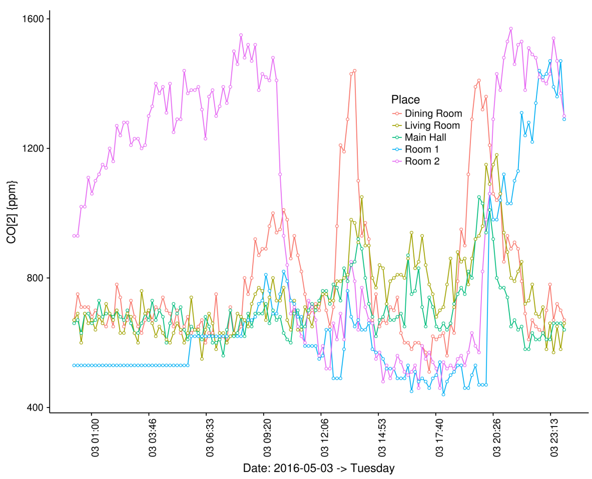

Fig. 5: CO_2 concentration levels (ppm), 3/5/2016.

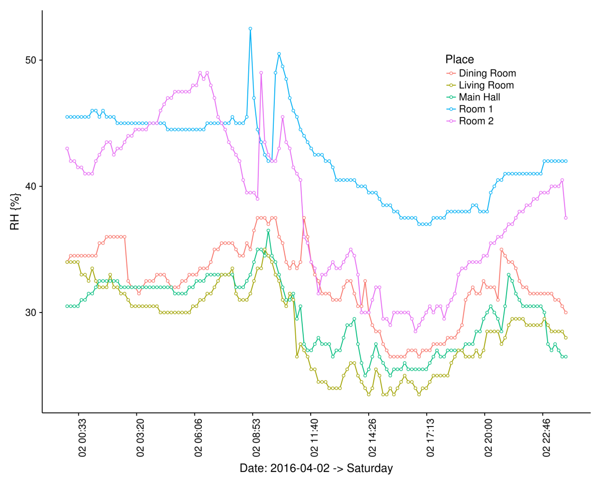

Fig. 7: Relative humidity values (%), 2/4/2016.

have been plotted dividing them by hour, which could indicate if the concentrations show similar patterns to the ones expected based on the time table. This way the patterns expected due to the time table and the controlled movements of the residents in the building can be analysed. Figures 8 and 9 show the patterns of the CO_2 measurements in May in the dining room and in the room 2.

The graphs show that as expected a increase in the CO_2 levels happens in the breakfast, meal and dinner times, while the rest of the day concentrations show similar values. A similar pattern can be observed in the room 2, in this case it is supposed to be occupied by night and empty during the day, that is exactly what the plot in Fig. 9 shows; with high concentration values during the night and with lower and nearly constant values during the day. It has to be mentioned that the sensor system in the rooms is quite near to the bed, which makes the CO_2 concentration to be such high.

To sum up, the CO_2 concentration shows a clear relation with the expected occupancy so it is supposed to have a great importance in the machine learning based models that have been built, those models are analysed in section 6. Not only the raw values seem to have importance, but the relative changes in the values should also be considered as the CO_2 levels in the rooms shown in figures 8 and 9 are quite different in the occupied intervals.

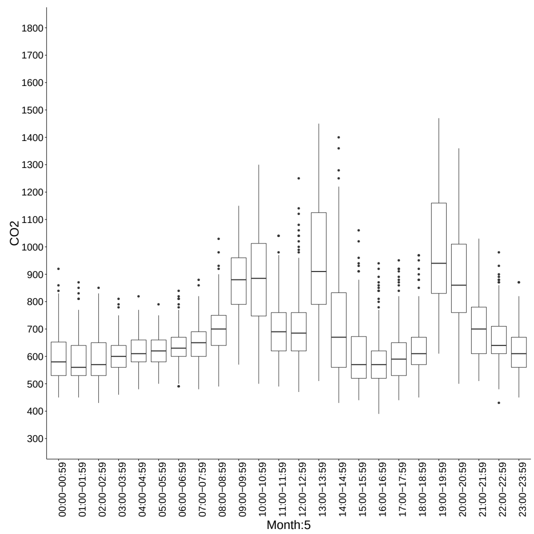

Fig. 8: CO_2 values per hour in the dining room on May.

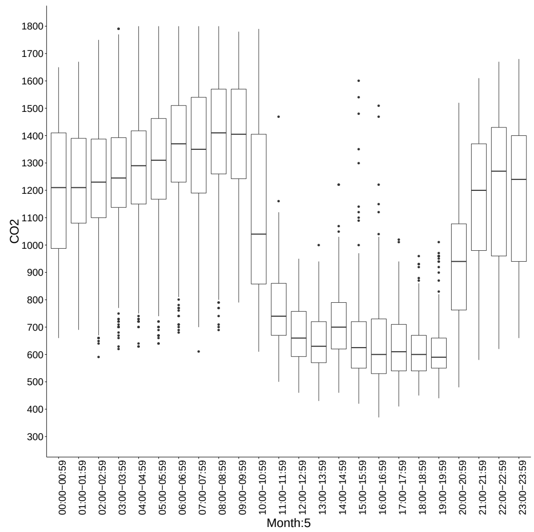

Fig. 9: CO_2 values per hour in the room 2 on May.

6 Results and Discussion

With the objective of determining occupancy based on data of indoor climate monitoring sensors, several classification models have been trained after preprocessing the data. The R project's language [22] was used for this purpose. A common organization of models was established by using the Caret package in R [23]. The models that were used to build the classifiers include classical classification trees, the gradient boosting method, multiclass Adaboost with bagging, C5.0 classification tree, Support Vector Machines, Quadratic Discriminant Analysis, and Neural Networks with one principal component analysis step. A ten-fold cross validation has been performed to validate the reliability of the models. The machine that was used to train the models was a Linux based SMP server with twenty cores running in parallel and 48 GB of available RAM.

As an example of the capability of the models that were trained, Fig. 6 shows the performance of different models, as well as accuracy and Cohen's kappa coefficient, which relates the obtained and expected accuracy of qualitative (categorical) items, on the case of the dining-room when validated against a dataset that was not previously seen during the training phase. It can be seen that, in this case, the model that performs better is the one that was based on supported vector machines. In general, it can be said that it performs noticeably better than the classical lineal models, ctree2 in the case of trees or the C5.0 rules. Bagging methods, such as random forest (RF), the extreme gradient boosting (xgbtree) and Adaboost with bagging (Adabag) also obtain remarkably accurate values. However, the model that was based on supported vector machines with a radial kernel performs slightly better. It obtains greater accuracy and Cohen's kappa coefficient values.

In order to have a better insight in the predictions of the models, Fig. 11 shows the ground truth in the dining room, the predictions done by the random forest classifier, the model which obtains highest accuracy values, and the differences between them for a given day. Some error do appear in the predictions, as the one after diner in this case, but in general the predictions are acceptable. Even if the models are not completely accurate, it can be seen how the predictions make sense, showing that binary occupancy can be determined by using indoor air quality data.

The most accurate models obtained with the set of classifiers that were trained, give accuracy values of about 80%. This is quite a good value, as some effects, such as the air flow between the rooms, have not been considered, and the predictions shown in the figure 11 do make sense. Furthermore, it is expected that the accuracy values will increase when taking those effects into account, which will be considered for further research activities.

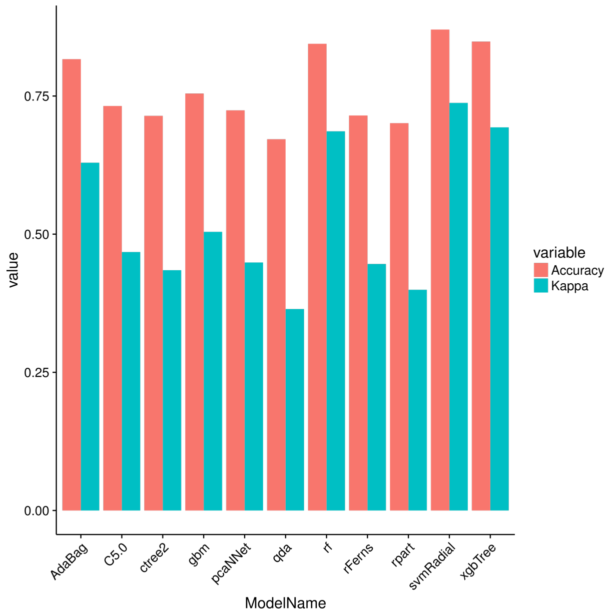

Fig. 10: Results obtained with the trained models.

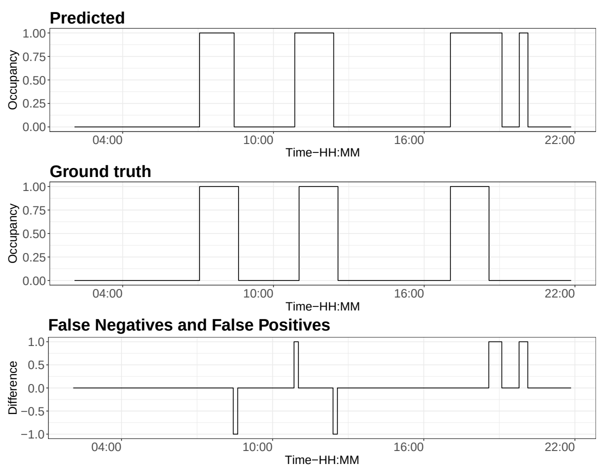

Fig. 11: Predictions, ground truth and differences with the RF model.

7 Conclusions

In general, the accuracy obtained by the models gives confidence in the identification of binary occupancy in the rooms. Even if the monitoring has not been performed for this particular purpose, we have shown that it is possible to gain additional information about the house usage. Nevertheless, there is place to improve, the confusion rate of the non-linear models is not big, but hopefully it could be reduced even more by performing a feature engineering. The main reasons for the obtained confusion rates are that the raw variables as their intrinsic variations only enforce basic rules, but reality is more complex.

As it has been mentioned through the paper, there are some relations and phenomena that have not been taken into account in this analysis, but they seem to have importance, or at least they are possible ways for getting errors. Those phenomena include the air flow between the adjoining rooms for example. The previously identified effects of the air flow through the rooms suggest including specific information about variable trends over time. They also suggest considering the building's semantic as a key element in order to learn from the different patterns and avoid false positives due to the air flow between adjoining rooms. It has also been seen that the opening of the windows with ventilating purposes changes considerably the humidity values, so it could be interesting to take that into account. A further study should take all those phenomena into account in order to achieve greater accuracy values.

All those improvements will be considered for further development and research in the near future, in order to reduce uncertainty and provide stronger foundations for the development of a decision making system framework. This research shows that the CO_2 concentration is valid to measure the occupancy of a room even in a real scenario, while it is remarkable the contribution made in the context of understanding the building scenario as a key element when working on real interconnected buildings.

References:

[1] J. Lloret, A. Canovas, S. Sendra, and L. Parra, "A smart communication architecture for ambient assisted living," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 26–33, 2015.

[2] 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.

[3] A. Kumar and G. P. Hancke, "Energy efficient environment monitoring system based on the IEEE 802.15. 4 standard for low cost requirements," *IEEE Sensors Journal*, vol. 14, no. 8, pp. 2557–2566, 2014.

[4] S. C. Folea and G. Mois, "A low-power wireless sensor for online ambient monitoring," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 742–749, 2015.

[5] S. Abraham and X. Li, "A Cost-Effective Wireless Sensor Network System for Indoor Air Quality Monitoring Applications," *Procedia Computer Science*, vol. 34, pp. 165–171, 2014.

[6] M. Kotol, A. Heller, and C. Orthmann, "Introduction of flexible monitoring equipment into the Greenlandic building sector," *ARTEK Event 2014*, 2014.

[7] I. Richardson, M. Thomson, and D. Infield, "A high-resolution domestic building occupancy model for energy demand simulations," *Energy and buildings*, vol. 40, no. 8, pp. 1560–1566, 2008.

[8] B. Dong and B. Andrews, "Sensor-based Occupancy Behavioral Pattern Recognition For Energy And Comfort Management In Intelligent Buildings," *Eleventh International IBPSA Conference*, pp. 1444–1451, 2009.

[9] S. K. Wang, S. P. Chew, M. T. Jusoh, A. Khairunissa, K. Y. Leong, and A. A. Azid, "WSN based indoor air quality monitoring in classrooms," in *AIP Conference Proceedings*, vol. 1808, p. 20063, AIP Publishing, 2017.

[10] G. de Gennaro, P. R. Dambruoso, A. D. Loiotile, A. Di Gilio, P. Giungato, M. Tutino, A. Marzocca, A. Mazzone, J. Palmisani, and F. Porcelli, "Indoor air quality in schools," *Environmental chemistry letters*, vol. 12, no. 4, pp. 467–482, 2014.

[11] T. Vehviläinen, H. Lindholm, H. Rintamäki, R. Pääkkönen, A. Hirvonen, O. Niemi, and J. Vinha, "High indoor CO2 concentrations in an office environment increases the transcutaneous CO2 level and sleepiness during cognitive work," *Journal of occupational and environmental hygiene*, vol. 13, no. 1, pp. 19–29, 2016.

[12] K. Tijani, S. Ploix, B. Haas, J. Dugdale, and Q. D. Ngo, "Dynamic Bayesian Networks to simulate occupant behaviours in office buildings related to indoor air quality," *arXiv preprint arXiv:1605.05966*, 2016.

[13] G. Diraco, A. Leone, and P. Siciliano, "People occupancy detection and profiling with 3D depth sensors for building energy management," *Energy and Buildings*, vol. 92, pp. 246–266, 2015.

[14] X. Guo, D. K. Tiller, G. P. Henze, and C. E. Waters, "The performance of occupancy-based lighting control systems: A review," *Lighting Research & Technology*, vol. 42, no. 4, pp. 415–431, 2010.

[15] M. Jin, N. Bekiaris-Liberis, K. Weekly, C. Spanos, and A. Bayen, "Sensing by proxy: Occupancy detection based on indoor CO2 concentration," *UBICOMM 2015*, p. 14, 2015.

[16] Z. Chen, Q. Zhu, M. Masood, and C. S. Yeng, "Environmental Sensors based Occupancy Estimation in Buildings via IHMM-MLR," *IEEE Transactions on Industrial Informatics*, 2017.

[17] D. Wörner, T. von Bomhard, M. Röschlin, and F. Wortmann, "Look twice: Uncover hidden information in room climate sensor data," in *Internet of Things (IOT), 2014 International Conference on the*, pp. 25–30, IEEE, 2014.

[18] M. Hassanalieragh, A. Page, T. Soyata, G. Sharma, M. Aktas, G. Mateos, B. Kantarci, and S. Andreescu, "Health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: Opportunities and challenges," in *Services Computing (SCC), 2015 IEEE International Conference on*, pp. 285–292, IEEE, 2015.

[19] I. Chiuchisan and O. Geman, "An approach of a decision support and home monitoring system for patients with neurological disorders using internet of things concepts," *WSEAS Transations on Systems*, vol. 13, pp. 460–469, 2014.

[20] ISO/IEC Standard 14543-3-10, "International standard ISO/IEC 14543-3-10: Wireless Short-Packet (WSP) protocol optimized for energy harvestingarchitecture and lower layer protocols," 2012.

[21] S. Noye, R. North, and D. Fisk, "Smart systems commissioning for energy efficient buildings," *Building Services Engineering Research and Technology*, vol. 37, no. 2, pp. 194–204, 2016.

[22] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[23] M. K. C. from Jed Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, and C. Candan., *caret: Classification and Regression Training*, 2016.