# An IoT−based system that aids learning from human behavior: A potential application for the care of the elderly

*Unai* Saralegui[1], *Miguel Ángel* Antón[1] and *Joaquín* Ordieres-Meré[2,*]

[1]TECNALIA Research & Innovation, Sustainable Construction Division, Area Anardi 5, 20730 Azpeitia, Gipuzkoa, Spain
[2]Universidad Politécnica de Madrid, ETSII, PMQ Research Group, José Gutierrez Abascal 2, 28006 Madrid, Spain

**Abstract.** The goal of this paper is to describe the way of taking advantage of the non-intrusive indoor air quality monitoring system by using data oriented modeling technologies to determine specific human behaviors. The specific goal is to determine when a human presence occurs in a specific room, while the objective is to extend the use of the existing indoor air quality monitoring system to provide a higher level aspect of the house usage. Different models have been trained by means of machine learning algorithms using the available temperature, relative humidity and $CO_2$ levels to determine binary occupation. The paper will discuss the overall acceptable quality provided by those classifiers when operating over new data not previously seen. Therefore, a recommendation on how to proceed is provided, as well as the confidence level regarding the new created knowledge. Such knowledge could bring additional opportunities in the care of the elderly for specific diseases that are usually accompanied by changes in patterns of behavior.

## 1 Introduction

To maintain the health of people, it can be useful to monitor their health status in their daily lives. Ambient assisted living (AAL), a term that first appeared in the European Framework Program for research funding, are systems that are intended to improve the quality of life of special groups of people, including the elderly. Such systems make use of information and communication technologies to develop applications and services for the elderly, in order to help them in their daily activities.

When implementing an AAL system, several features should be considered [1]. First, the sensors should be non-invasive systems. They can be embedded in wearable items, like shoes, etc., or can be part of the infrastructure. In any event, it is important that these systems respect the privacy of the users.

Elderly people who live alone usually suffer more accidents than other people. For example, in the district of Kaiserslautern, Germany, 30 per cent of persons who are more than 65 years of age and living 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 the existing infrastructure, will help in providing specific care to such a population segment without a need for additional investment.

The intention of this paper is to analyze the opportunity to pursue additional goals and obtain further and relevant information by means of an already deployed infrastructure, in order to bring higher value to the user, in connection with the Internet of Things (IoT) paradigm.

As we seek to monitor the behavior of the elderly to detect changes in their patterns of behavior, 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 use of the sensors that are already there to monitor the air quality.

A network of wirelessly connected sensors is used to obtain the data from rooms in an elderly caring institution. Such sensors are designed to monitor the air quality and measure temperature, relative humidity and $CO_2$ concentrations. The data is analyzed and some models are trained with the data to detect the presence of a person in a room. This requires machine learning algorithms.

In order to better understand the work being done, the structure of the paper will first introduce the problem by analyzing the present state of the art in the following section. The methodology that is adopted for this on-going research will be presented in the third Section. The fourth section will provide the selected study case, the main assumptions made in this work and details of the sensor type and implementation. Section five will include the discussion and the analysis of the findings. Finally, the last section will give the main conclusions of the work and discuss future developments and research.

## 2 State of the Art

There are several paths for research of the use of IoT technology to acquire greater knowledge of different behaviors. In this area, studies of air quality and energy

---

Corresponding author: j.ordieres@upm.es

efficiency in buildings have received much attention as in [2-3] where indoor monitoring is proposed and occupancy is related to comfort measurements.

Another interesting trend involves monitoring the indoor air quality, 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.

Not only has energy efficiency or comfort−related topics been analyzed, but also IoT technologies have been used to monitor pollutant levels in indoor areas. A correlation between the concentration of pollutants and health problems in schoolchildren has also been shown in some studies [7-8]. 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 [9-10].

An area that is receiving increasing attention is the demand-driven HVAC (Heat Ventilation and Air Conditioning) control, for both energy efficiency and indoor comfort. This is seen in the case of [11], where the estimation of a building's occupancy by use of a wireless $CO_2$ sensor network is investigated. The HVAC system can take advantage of the information, acquired about non-occupancy, in order to save energy and resources.

Another trend that is more closely related to our research interest involves using an existing air quality monitoring system to derive occupancy information in a non-intrusive scenario. The key aspect is to work under long term service principle. Various approaches have been examined to determine the occupancy information as video-image based detection [12] or detection by passive infrared (PIR), sensors [13]. However, both approaches have limitations. These involve privacy or computational cost in the case of video, and a lack of movement in the case of PIR sensors. Those limitations have caused $CO_2$−based presence detection methods to receive increasing attention.

In our area of research interest, mixed approaches, such as [14], 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) [15]. Some research to determine binary occupancy information also has been undertaken, as in [16] 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, as the sampling

system cannot be reconfigured from its primary application and somehow is limited in its operating life.

# 3 Methodology

The adopted framework relies on the existing infrastructure for air quality monitoring.

In addition, the data is collected and uploaded to the cloud. Therefore, data can be accessed from any place and models for derivation of the behavior rules can be implemented.

The framework will use as many sensors as are available, as well as the most convenient techniques for data processing and modeling based on machine learning techniques. In this first step, we will analyze the data that is obtained and determine possible relationships that may help in building the monitoring framework. When sufficient knowledge is available and the desired information can be obtained, the decision-making processes will be applied, depending on the case of interest. (See Fig. 1.)
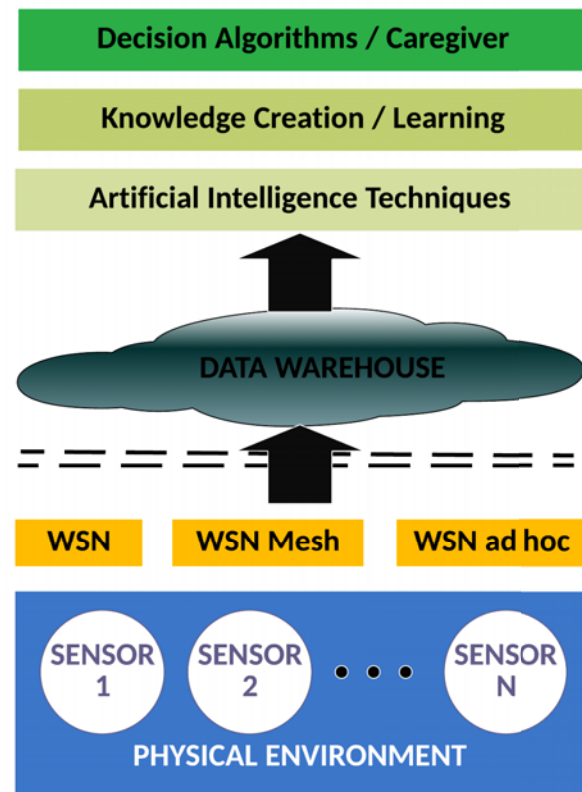


**Fig. 1.** The framework adopted for the research.

It is worth remembering that the main goal here is to determine human presence in specific areas as an example of a higher level of information. Our main hypothesis in this phase of the study is that it is possible to determine the binary occupancy based on the local data available at room level. As the goal is to use an already available infrastructure, the monitoring frequency is kept at the same value as in its initial application. That is a sampling frequency of 10 minutes, which means that the system is not sensitive to high

frequency movements. However, it makes it possible to maintain a low energy consumption.

It will be also interesting in this research to determine the best set of algorithms to predict the occupancy values, in order to obtain the model can provide the best results for this problem. To this end, the problem is treated as a supervised learning process, in which linear and nonlinear techniques will be used with the models, in a first approach that is being built for each monitored room.

# 4 Study case

An agreement with a caring institution for the elderly has permitted us to collect the data that is necessary for this study. The institution has allowed us to install and monitor sensors in five different rooms in the building. The monitored rooms include the main hall, the living room, the dining room and two bedrooms.

The monitoring period lasted for one year and respected the national laws regarding non-intrusiveness and confidentiality. Five identical wireless sensors were placed in the rooms to measure and report levels of ambient temperature (ºC), relative humidity (%) and $CO_2$ concentration (ppm). The monitoring frequency was set at ten minutes to maintain the energy requirements.

During normal operation, the sensors are powered by ambient room light by means of an energy harvesting solar cell. In prolonged periods of low light, security backup batteries provide power to the sensors. The sensors support the open EnOcean® standard protocol (ISO/IEC 14543-3-10) for wireless communications [17]. The data is transmitted wirelessly to a central node that uploads it to the cloud.

Doors separate the rooms in the monitored building. While the data was being collected, thirteen residents were living in the building's monitored rooms. Because of the condition of the residents, their routines are well−defined and maintained throughout the year. Thus, abnormalities are easy to measure. 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.

# 5 Results and Discussion

By performing a visualization of the data for each room that each sensor generates, interesting relationships are suggested. From the brief overview in Fig. 2 it becomes clear that the dining-room is used mainly used at around 9:00 a.m., 1:00 p.m. and 8:00 p.m., which agrees with the scheduled times for breakfast, lunch and dinner in the institution.

The effect of the doors being opened or closed has also been visualized. As can be seen in the line that represents Room 2 in Fig. 2, the $CO_2$ concentration levels increase continually, whereas the concentration in Room 1 remains more stable and even decreases in some phases. Such behavior is expected in a room that has an

open door, as the air flows through rooms whose entrances present no barrier.

The evolution of the temperature values appears in Fig. 3. The evolution is quite stable over time as the temperature values do not change much. This indicates that they will not provide sufficient information to detect changes in the occupancy of a room.

The data visualization enables us to see the impact on adjacent spaces when the doors remain open and the air flows from room to room. It becomes possible to visualize the effects as ventilation in the values of relative humidity around 9:00 a.m. in both rooms, see Fig. 4, showing a noticeable change in the relative humidity values at those moments. Consequently, the visualization of data shows that it is possible to detect when a window is open, allowing to monitor such activity.

In Fig. 5, a straight line with no change is observed during the night for Room 1. This means that there has been a problem in communication and the last measured value is provided.

By visualizing the time series that the sensor values describe, it appears that, when the doors are opened, the air flows between the rooms making the concentration of $CO_2$ to create different patterns. That is why it will be interesting in further research to add additional information to model the air flow between the rooms.
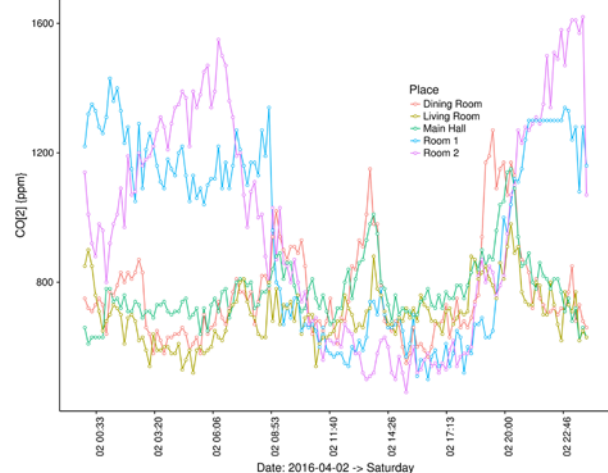


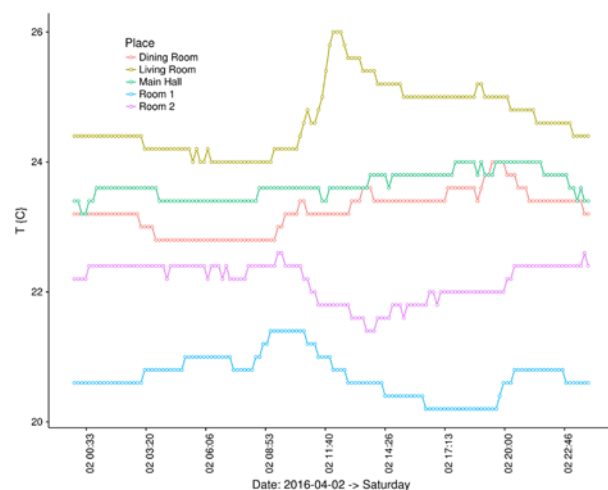**Fig. 2.** Evolution of the $CO_2$ concentrations, 2/4/2016.



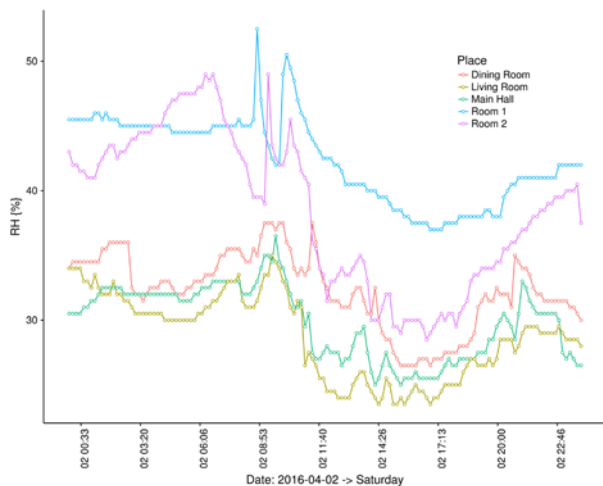**Fig. 3.** Evolution of the Temperature, 2/4/2016.

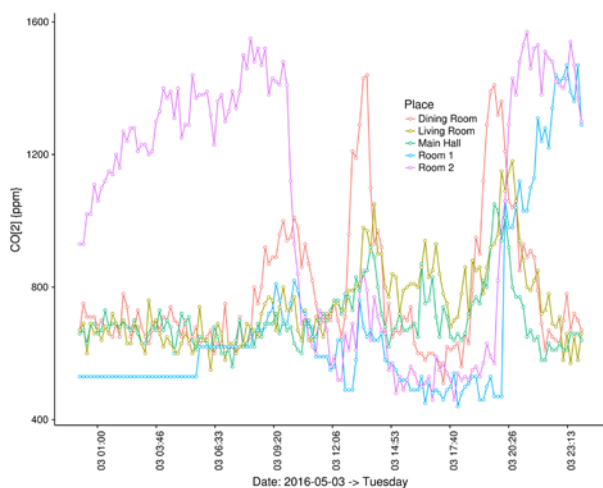**Fig. 4.** Evolution of the Relative Humidity, 2/4/2016.



**Fig. 5.** Evolution of the $CO_2$ concentrations, 3/5/2016.

As mentioned, the main objective in this phase of the study is to determine occupancy based on the data of an already available infrastructure in a real scenario. For this purpose, several classification models have been trained after pre-processing the data. The R project's language [18] was used for this purpose. A common organization of models was established by using the Caret package in R [19]. The models that were used to build the classifiers include classical classification trees, the gradient boosting method, the multiclass Adaboost with bagging, the C5.0 classification tree, the Support Vector Machines, Quadratic Discriminant Analysis, and Neural Networks with one principal component analysis step. 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.

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. It is expected that the accuracy values will increase when taking those effects into account, which is considered for further research activities.
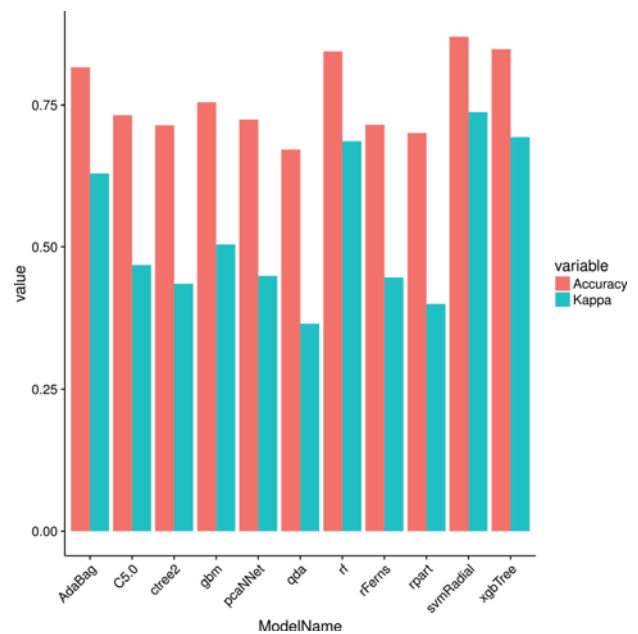


**Fig. 6.** Cohen's kappa coefficient and accuracy for the set of models that were trained.

# 6 Conclusions

The accuracy obtained by the models that were built gives confidence in the identification of binary occupancy in the rooms. However, there is room for improvement in reducing the confusion rate, although it has been reduced already in the case of non-linear classifiers. The main reasons for the confusion are that the bare variables and their intrinsic variations only enforce basic rules, but reality is more complex. For instance, indoor climate control enforces temperature within a short range. However, differences occur when the residents breathe with the room door open or closed. Actually, when the door remains open, the $CO_2$ levels and humidity spread to the adjoining areas and cause a rise in the concentration of such values in those areas. Of course, this effect does not occur in the same way when the doors are closed as the concentrations only rise in rooms that are occupied.

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.

All of 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.

## References

[1] J. Lloret, A. Canovas, S. Sendra, L. Parra, *A smart communication architecture for ambient assisted living*, IEEE Commun. Mag., **53(1)**, p. 26–33 (2015)

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

[3] A. Kumar, G. P. Hancke, *An energy-efficient smart comfort sensing system based on the IEEE 1451 standard for green buildings*, IEEE Sens. J., **14(12)**, p. 4245–4252 (2014)

[4] S. C. Folea, G. Mois, *A low-power wireless sensor for online ambient monitoring*, IEEE Sens. J., **15(2)**, p. 742–749 (2015)

[5] S. Abraham, X. Li, *A cost-effective wireless sensor network system for indoor air quality monitoring applications*, Procedia Comput. Sci., **34**, p. 165–171 (2014)

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

[7] S. K. Wang, S. P. Chew, M. T. Jusoh, A. Khairunissa, K. Y. Leong, A. A. Azid, *WSN based indoor air quality monitoring in classrooms*, AIP Conference Proceedings, **1808(1)**, p. 20063 (2017)

[8] G. de Gennaro, P. R. Dambruoso, A. D. Loiotile, A. Di Gilio, P. Giungato, M. Tutino, A. Marzocca, A. Mazzone, J. Palmisani, F. Porcelli, *Indoor air quality in schools*, Environ. Chem. Lett., **12(4)**, p. 467–482 (2014)

[9] T. Vehviläinen, H. Lindholm, H. Rintamäki, R. Pääkkönen, A. Hirvonen, O. Niemi, J. Vinha, *High indoor CO2 concentrations in an office environment increases the transcutaneous CO2 level and sleepiness during cognitive work*, J. Occup. Environ. Hyg. **13(1)**, p. 19–29 (2016)

[10] K. Tijani, S. Ploix, B. Haas, J. Dugdale, Q. D. Ngo, *Dynamic Bayesian Networks to simulate occupant behaviours in office buildings related to indoor air quality*, arXiv Prepr. arXiv1605.05966 (2016)

[11] Q. Huang, C. Mao, *Occupancy estimation in smart building using hybrid CO2/light wireless sensor network*, J. Appl. Sci. Arts, **1(2)**, p. 5 (2017)

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

[13] X. Guo, D. K. Tiller, G. P. Henze, C.E. Waters, *The performance of occupancy-based lighting control systems: A review*, Lighting Research & Technology, **42(4)**, p. 415–431 (2010)

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

[15] Z. Chen, Q. Zhu, M. Masood, C. S. Yeng, *Environmental Sensors based Occupancy Estimation in Buildings via IHMM-MLR*, IEEE Trans. Ind. Informatics (2017)

[16] D. Wörner, T. von Bomhard, M. Röschlin, F. Wortmann, *Look twice: Uncover hidden information in room climate sensor data*, 2014 International Conference on the Internet of Things, p. 25-30 (2014)

[17] ISO/IEC Standard 14543-3-10, *International standard ISO/IEC 14543-3-10: Wireless Short-Packet (WSP) protocol optimized for energy harvesting— architecture and lower layer protocols*, edition 1.0 (2012)

[18] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria (2016)

[19] M. Kuhn 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, C. Candan, *caret: Classification and Regression Training* (2016)