ISSN 1870-4069

Environmental Monitoring for Estimation of Indoor Occupancy Levels based on Unsupervised Machine Learning for COVID-19 Restrictions

Alma Rosa Mena Martínez, Héctor Ceballos,
Joanna Alvarado-Uribe

Instituto Tecnológico y de Estudios Superiores de Monterrey,
Mexico

a00834070@tec.mx,ceballos@tec.mx,joanna.alvarado@tec.mx

Abstract. Today, the COVID-19 pandemic has imposed stricter regulations on social distancing and crowd control, these policies can be managed effectively through the occupancy information of the place. Nevertheless, to predict the number of occupants accurately, there needs to be sufficient labeled data for training the predictors and validating the model. Furthermore, how the information is collected can represent a problem because direct methods such as cameras invade the personal privacy. Therefore, the main purpose of this research is the development of an affordable and non-intrusive low-cost solution to improve the management of public spaces and ensure the compliance of government regulations regarding the number of occupants inside, implementing unsupervised and semi-supervised machine learning models as well as data fusion from different Internet of Things (IoT) environmental sensor nodes. The preliminary results show that the clusters follow a pattern during the day, therefore this can be used to assign the occupancy levels.

Keywords: Environmental sensors, machine learning, data fusion, occupancy estimation.

1 Introduction

Today, we are facing an unprecedented world health problem derivative from SARS-CoV-2 which has changed all the routines and habits around the world. Preventive measures have been implemented to allow economic and social activities to be carried out while minimizing the impact on people's health. Such measures include the use of face masks, physical distancing, bans on large gatherings, mobility restrictions, among others [4]. It is important to flag out, that even after you're vaccinated, you should keep taking precautions [9].

Therefore, the main purpose of this research is the development of an affordable and non-intrusive low-cost solution to improve the management of public spaces and ensure the compliance of government regulations regarding the number of occupants inside. The proposed solution monitors air temperature, relative humidity, and barometric pressure and apply unsupervised and semi-supervised machine learning algorithms as well as fusion data techniques to estimate the occupancy levels in enclosed spaces. The proposed device to collect the data is equipped with a micro-controller ESP32 [13], and one BME280 [12] sensor.

This paper is structured as follows. Section 2 discusses the related work. Section 3 presents the Hypothesis and Research Questions. Section 4 shows the research objectives. Section 5 presents the methodology to follow. Finally, the state of the research and conclusions are presented in Section 6 and 7.

2 Related Work

During the past years, several research focused on occupancy detection and estimation had been carried out intending to reduce energy consumption, maximizing comfort, security management, and so on [6, 19, 17].

Solutions for COVID-19 Restrictions

Some researches have proposed solutions to handle the COVID-19 preventive measure. For instance, Longo et al. [7] presents a prototype "Smart Gate", whose main function is to perform people flow monitoring and to keep track of the occupancy levels of both indoor and outdoor spaces. Floris et al. [5] proposes an IoT-based smart building solution for indoor environment management, which aims to provide main functionalities as monitoring environmental parameters of the room, detection of the number of occupants in the room, among others.

Design and Development of an Occupancy Sensor

Researchers from The Tecnologico de Monterrey designed a device which estimating the occupancy levels in closed spaces. The device is equipped with a micro-controller ESP32 [13], a BME280 [12], and other electronic components. The controller is programmed to measure every minute the temperature, barometric pressure, and humidity. The data is sent through Wi-Fi to store it in a data-based cloud. Vela et al. [15] carried out occupancy level estimation research, deploying the proposed device within a university gym and in a living room. The accuracy obtained was around 95.2% to 97% using Support Vector Machine (SVM), K-nearest neighbor (k-NN), and decision tree (DT) models.

3 Hypothesis and Research Questions

It is possible to approximate with an accuracy at least of 90% the occupancy levels in enclosed space using unlabeled fusion data from environmental sensing such as air temperature, relative humidity, and atmospheric pressure. Some *research questions* are:

– How many sensors are necessary to install within a room depending on its dimension?
– How deploy the sensors? (position,height,etc.)

– Is essential to add another environmental variable or non-intrusive sensor to estimate occupancy levels in real-time with an accuracy at least of 90%?
– Is it possible to use unlabeled data for estimating the occupancy levels?
– Could data fusion improve the performance of the unsupervised machine learning model?
– Could externals weather conditions affect the achievement of the accurate estimation? Due to the Covid-19 pandemic in which it is necessary to have natural ventilation (windows and/or door open)

4 Objectives

The general objective of this work is to apply unsupervised and semi-supervised machine learning algorithms as well as data fusion methods to estimate the occupancy levels indoors using only ambient variables. The particular goals be achieve as this research work is conducted are:

– Select three possible scenarios with different design-use and dimensions.
– Set the number of nodes (microcontroller with the sensors) necessaries to deploy inside the place.
– Collect data without labels and data with labels.
– Process the datasets and implement a data fusion method.
– Develop an unsupervised and semi-supervised machine learning algorithm.
– Assess the performance of the models and the number of sensors installed by a square meter.
– Analyze whether natural ventilation through doors and windows affects to the machine learning models performance.

Research Contributions The main contribution of this research will be an unsupervised and semi-supervised system to estimate the occupancy levels indoors, as well as the design of a data fusion framework, and the generation of three robust data sets collected in enclosed spaces which will be available to everyone for future analysis.

5 Methodology

The main activities are described below:

1. ***Selection of the test-bed scenarios:*** Three different places with a specific design use will be selected. For instance, meeting rooms, multi-occupant office, and collaborative spaces.
2. ***Sensor Deployment:*** To set the number of devices deployed within each space will be based on the place's superficial area and the literature review.
3. ***Data collection:*** For 3 weeks, an unlabeled data will be collected.
4. ***Data pre-processing:*** It is essential to seek a missing value and outliers, erroneous measures, remove the noise, and the extraction of features.

5. ***Data Exploration:*** Visualizations such as heat maps, histograms, and 3D scatter plots will be carried out. The main goal is to find patterns and hidden substructure which helps us to interpret the outcomes from the models.
6. ***Data Fusion:*** In this stage, the aim is to archive a better performance of the models, integrating the data collected from multiple sensors.
7. ***Machine Learning model selection:*** Through the literature review, the semi-supervised and the unsupervised algorithms will be selected. The total of clusters and/or classes desired is four that will be identified as empty, low, medium, and high occupancy levels.
8. ***Model evaluation:*** The Internal validation index established the quality of the clustering structure without having access to external information [10]. Nevertheless, the clusters need to be compared with the real occupancy. Evaluating it with metrics as Accuracy, and confusion matrix. Therefore, metrics such as accuracy and confusion matrix will be used.

6 State of the Research

During the year 2021, the research work was focused on improving the proof-of-concept system designed by the peers, it is essential because the next experiments involve the deployment of several devices *in-situ*. Furthermore, preliminary experiments were carried out; first, the k-means algorithm was fitted using the data collected by Vela et al. [15], to verify that *"unlabeled"* data can be used to estimate occupancy levels. Second, the prototype was placed in a multi-occupant office belongs to Tecnologico de Monterrey. The data were collected for two weeks to develop unsupervised algorithms such as k-means, agglomerative, and Fuzzy C-means. Nowadays, the research work will start with the selection of the three test-bed enclosed spaces, the sensor deployment, the collection and analysis of data.

7 Preliminary Results

The preliminary results obtained during February to September, 2021 are presented.

Prototype as a Minimum Viable Product (MVP).

The prototype is currently an MVP. Its components were designed and printed as a printing circuit board (PCB) of 5 cm x 5 cm, covered by an external case (6.5 cm x 6.5 cm x 3 cm). This prototype can be easily installed anywhere in the enclosed space.

K-means for living room data.

From the data, the classes was deleted; then the data were normalized and standardized. Four clusters were set to seek. The results show similarities with the classes from the original data. In other words, the clusters formed can be used as classes (empty, low, mid, high) which indicate the occupancy levels.

K-means, agglomerative and Fuzzy C-means for office data.

The data collected in the multi-occupant office do not have any label, the attendance record was requested to verify the real occupancy; therefore the algorithms had not yet been evaluated as predictors of the occupancy levels. Nevertheless, the evaluation of clusters using Silhouette Index was 60% for k-means, 64% for agglomerative, and 67% for Fuzzy C-means. The cluster obtained follow the patters of the temperature and humidity during the day.

References

1. Aosong Electronics Co. Ltd.: DHT22 (DHT22 also named as AM2302). https://www.sparkfun.com/datasheets/Sensors/Temperature/DHT22.pdf
2. Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W.C., Wang, C.B., Bernardini,S.: The Covid-19 pandemic. Critical Reviews in Clinical Laboratory Sciences 57(6), 365–388 (2020). https://doi.org/10.1080/10408363.2020.1783198, https://doi.org/10.1080/10408363.2020.1783198, pMID: 32645276
3. Fajilla, G., Austin, M.C., Mora, D., Simone, M.D.: Assessment of probabilistic models to estimate the occupancy state in office buildings using indoor parameters and user-related variables. Energy and Buildings 246, 111105 (sep 2021). https://doi.org/10.1016/j.enbuild.2021.111105
4. Fernandez-Carames, T.M.; Froiz-Miguez: An IoT and Blockchain based System for Monitoring and Tracking Real-time Occupancy for COVID-19 Public Safety. In: Proceedings of the 7th International Electronic Conference on Sensors and Applications. Switzerland (Nov 2020). https://doi.org/https://doi.org/10.3390/ecsa-7-08207
5. Floris, A., Porcu, S., Girau, R., Atzori, L.: An IoT-Based Smart Building Solution for Indoor Environment Management and Occupants Prediction. Energies 14(10), 2959 (may 2021). https://doi.org/10.3390/en14102959
6. Han, K., Zhang, J.: Energy-saving building system integration with a smart and low-cost sensing/control network for sustainable and healthy living environments: Demonstration case study. Energy and Buildings 214, 109861 (may 2020). https://doi.org/10.1016/j.enbuild.2020.109861
7. Longo, E., Redondi, A.E.C., Bianchini, M., Bolzan, P., Maffei, S.: Smart gate: a modular system for occupancy and environmental monitoring of spaces. In: 2020 5th International Conference on Smart and Sustainable Technologies (SpliTech). IEEE (sep 2020). https://doi.org/10.23919/splitech49282.2020.9243754
8. Secretaria de Relaciones Exteriores de Mexico: Mexico Covid-19 monitoring system. (2021) https://embamex.sre.gob.mx/eua/index.php/en/2016-04-09-20-40-51/tourism/1760-mexico-s-covid-19-monitoring-system
9. World Health Organization: Covid-19 advice for the public: Getting vaccinated. (Sep 2021) https://www.who.int/ emergencies/ diseases/ novel-coronavirus -2019/ covid-19-vaccines/ advice
10. Palacio Nino, J.: Evaluation metrics for unsupervised learning algorithms. (2019) https://doi.org/arXiv:1905.05667
11. Patel, A. A.: Hands-On Unsupervised Learning Using Python. Orelly Media,Inc, first edition edn. (Feb 2019)
12. Bosch Sensortec: BME280 Digital humidity, pressure and temperature sensor. Bosch Sensortec, bst-bme280-ds002-15 edn. (Sep 2018), https://www.mouser.com/datasheet/2/783/BST-BME280-DS002-1509607.pdf

13. Espressif Systems: ESP32 Series. (2021) https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf
14. Trivedi, D., Badarla, V.: Occupancy detection systems for indoor environments: A survey of approaches and methods. Indoor and Built Environment 29(8), 1053–1069 (2020). https://doi.org/10.1177/1420326X19875621, https://doi.org/10.1177/1420326X19875621
15. Vela, A., Alvarado-Uribe, J., Davila, M., Hernandez-Gress, N., Ceballos, H.G.: Estimating Occupancy Levels in Enclosed Spaces Using Environmental Variables: A Fitness Gym and Living Room as Evaluation Scenarios. Sensors 20(22), 6579 (nov 2020). https://doi.org/10.3390/s20226579
16. Wei, Y., Xia, L., Pan, S., Wu, J., Zhang, X., Han, M., Zhang, W., Xie, J., Li, Q.: Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks. Applied Energy 240, 276–294 (apr 2019). https://doi.org/10.1016/j.apenergy.2019.02.056
17. Yang, Z., Li, N., Becerik-Gerber, B., Orosz, M.: A Multi-Sensor Based Occupancy Estimation Model for Supporting Demand Driven HVAC Operations. In: Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design. SimAUD '12, Society for Computer Simulation International, San Diego, CA, USA (2012)
18. Zheng, A.: Evaluating Machine Learning Models. O'Reilly Media, Inc, first edition edn. (Sep 2015)
19. Zimmermann, L., Weigel, R., Fischer, G.: Fusion of nonintrusive environmental sensors for occupancy detection in smart homes. IEEE Internet of Things Journal 5(4), 2343–2352 (aug 2018). https://doi.org/10.1109/jiot.2017.2752134