2020

Sensor Data Analysis in Smart Buildings

Manuel A. Mane Penton
CUNY New York City College of Technology

Sensor Data Analysis in Smart Buildings

Manuel A. Mane Penton

Advisor: Prof. Li Geng

Department of Electrical and Telecommunications Engineering Technology

Abstract

Data analysis and Machine Learning are destined to evolve the current technology infrastructure by solving technology and economy demands present mainly in developed cities like New York. This research proposes a machine learning (ML) based solution to alleviate one of the main issues that big buildings such as CUNY campuses have, that is the waste of energy resources. The analysis of data coming from the readings of different deployed sensors such as CO2, humidity and temperature can be used to estimate occupancy in a specific room and buildings in general. The outcome of this research stablished a Machine Learning model that could predict occupancy based in the values of CO2, temperature and humidity of the room.

Detection Occupancy in Smart Buildings

- Many researchers struggle to build error-free detection occupancy models taking into consideration factors such as cost, privacy and efficiency.
- These systems require continuous and mathematically challenging analysis of data coming from a series of sensors that need to be properly and exquisitely deployed in the room in order to provide the cleanest and most accuracy possible data.
- Even though they have come with different types of models and solutions, they present their own advantages and disadvantages depending mainly in the structure of the area and budget, as presented in the following comparison table:

Model	Description	Advantages	Disadvantages
Smarthphone - based	Use the bluetooth and WiFi signals coming from smartphones.	High accuracy	It requires the user to carry a smarthphone and to be connected to the local WiFi or having the Bluetooth on.
Dopple - Radar	Use the transmitted and received signals of dopple radars to detect vital signs	High accuracy	High cost, poor scalability and low efficient with obstacles.
Cameras - PIR	Use PIR (passive infra-red) sensors to detect movement, face and obstacles recognition.	High accuracy	PIR sensor need movemet to detect occupancy. Face and obstacles recognition require higher costs.
Environmental Features	Use sensor readings such as CO2, temperature, humidity, light and others.	Low cost, high scalability	Lower accuracy, it depends on the accuracy of the model.

- This research paper describes the Machine Learning modelling of an Occupancy Detention method based in CO2, temperature and humidity data.

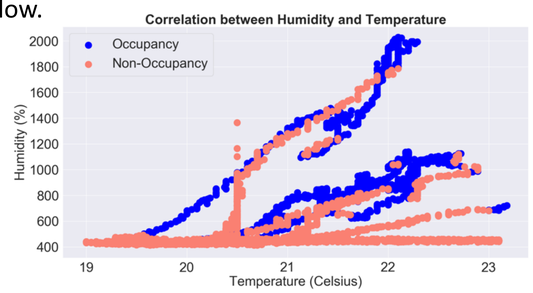
Methods

- This ML model is built using an experimental Occupancy Detection Data Set from UCI Machine Learning Repository as reference.
- The data is divided in one data set to train the model and two data sets to test it.
- They describe the values of Temperature (in Celsius), relative Humidity (%), and CO2 (in ppm). Our training data set contains 8143 rows and looks like the following:

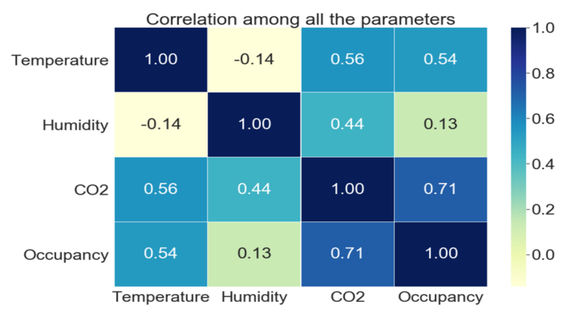
date	Temperature	Humidity	Light	CO2	HumidityRatio	Occupancy
2015-02-02 14:19:00	23.7000	26.272	585.200000	749.200000	0.004764	1

Procedures

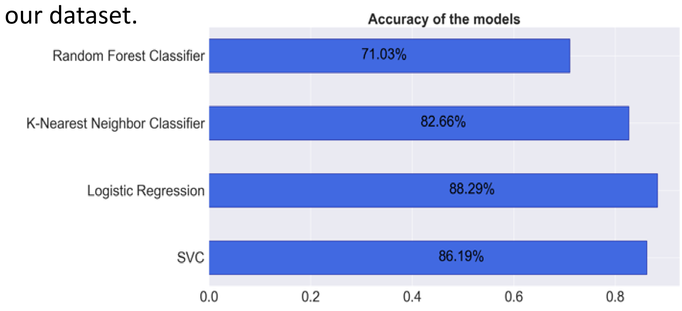
- A pre-analysis of our dataset allows us to get more information about the parameters and such as the correlation of the humidity and temperature in the situations of occupancy and non-occupancy shown below.



- The "Correlation Matrix" below shows the correlation of the variables with the occupancy. As result we noticed that CO2 and Temperature, in a lower rate, are the ones that more impact have when detecting occupancy. They both show a positive correlation, meaning they are directly proportional with Occupancy.

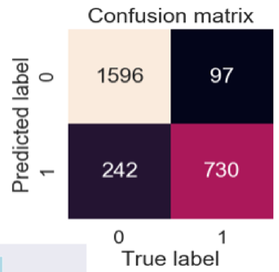


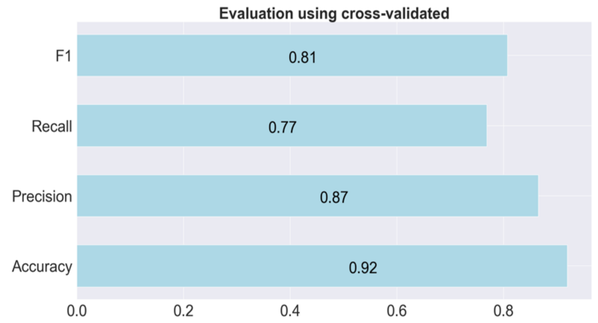
- After the analysis of the data variables, we evaluate different machine learning models to look for the one highest accuracy for our dataset.



Results

- The Logistic Regression has the highest accuracy, so, it is the most fitted model for this data.
- This model can still be improved by tuning its parameters. A deeper evaluation of the model also offers us more information of its efficiency for our data scenario.
- The "Confusion Matrix" display a comparison of the predicted values of the model and the real ones using the test dataset. At the same time, the "cross-validated" technique provides a deeper evaluation of our model.





Discussion

Depending of the requirements of the user, the accuracy of this model may not be the best. That's the biggest flaw of Detection Occupancy using environmental features as described before. However, the accuracy can be improved by collecting more testing data or adding another component to the data such as Light. The article "How to Predict Room Occupancy Based on Environmental Factors", Jason Brownlee (2018), describes a Machine Learning model that achieve a 99% accuracy in detecting occupancy by using similar data but adding more samples and variables.

Even though the accuracy is lower compared to other Detection Occupancy, this model is cheaper and can be adapted with other types of sensors such as PIR to increase its accuracy.

References

1. Occupancy Detection Data Set, UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/Occupancy+Detection
2. Jason Brownlee, "How to Predict Room Occupancy Based on Environmental Factors". 29 August 2018.
3. Bruno Abade, David Perez Abreu and Marilia Curado, " A Non-Intrusive Approach for Indoor Occupancy Detection in Smart Environments". *Science Direct.* 15 Nov. 2018.
4. Zhenghua Chen, Chaoyang Jiang and Lihua Xie, "Building occupancy estimation and detection: A review". *Science Direct.* 5 April, 2018.
5. H. Elkhoukhi, Y. NaitMalek, A. Berouine, M. Bakhouya, D. Elouadghiri and M. Essaaidi, "Towards a Real-time Occupancy Detection Approach for Smart Buildings". *Science Direct. 2018.*