Sensor Occupancy Detection Using XG Boost Algorithm

[1]G. Mounika, [2]S. Sai Kumar, [3]Dr. P. K. Sahoo, [4]K. Eswaran
[1,2,3,4]Dept. of CSE, SNIST, Hyderabad, Telangana, India

Abstract
A room in a smart home is fixed with environmental sensors for sensing of the indoor air quality. Environmental sensors can be any sensor from simple air temperature sensor to an indoor air quality measurement system, which holds different types of sensors or a networked sensor. Purpose of these sensors is to determine the indoor air quality and their potential in incorporating occupancy detection is largely unused. Occupancy detection is a technique used to detect the presence of living and non-living things. There are many environmental sensors which are used to detect different kinds of gases, namely CO_2 (carbon dioxide) and TVOC (total volatile organic compounds) sensors which are used here to detect the gases that resides in a room. By detecting the indoor gases we can improve the quality of the air. CO_2 sensor is used for detection of carbon dioxide composition, where as TVOC is internally built with CO_2 sensor and it will detect other gases too. There are many machine learning algorithms that are used to classify the occupancy detection. In previous studies, naive Bayes classifier is used for detecting occupants using Weka tool. Now In this paper XGBoost, a machine learning algorithm is used for detecting occupants.

Keywords
XGBoost, Environmental Sensors.

I. Introduction

Occupancy detection is used to detect the indoor gases in smart buildings. By detecting the indoor gases we can improve the air quality. There are many sensors in the real world, by using those sensors we can reduce the man power. We will place the sensors inside the closed room, they automatically sense the gases which are inside the room based upon the type of sensor. Different kinds of sensors are used for different applications. There are temperature sensors which will automate the room temperatures based on heat and cold conditions, proximity sensors are used for detecting an object in the room, IR Sensor (Infrared Sensor) is like proximity sensor and which is mainly used in transmission and reflection of light, and there are few other sensors namely pressure sensor, light sensor, ultrasonic sensor, smoke gas and alcohol sensor, touch sensor, color sensor, humidity sensor, tilt sensor, flow level sensor [1].

The goal of smart buildings is to reduce energy consumption and provide healthy indoor environments. Insufficient indoor air quality leads to poor and less ventilation. In many organizations, dizziness building syndrome may be a reason, and may result in HVAC systems that do not effectively distribute air to occupants. Homes with plenty of air filtration, the conventional filtration mechanism within the HVAC system will filter the air by removing the pollutants. If the outdoor air quality is poor (such as in major cities), this case also effects purification. For closest buildings, potency mechanical ventilation systems will guarantee clean indoor air. Efficient HVAC systems will make sure that occupants get adequate, filtered, fresh air. They ought to deliver the proper quantity of air to the required locations. Typically, inefficient systems aren't solely energy-inefficient, however additionally ineffective at doing their meant perform [7].

Business square measure is more and more curious about air quality observation systems to maintain the conditions that maximize employee productivity and save energy. To fulfill this, TSI manufactures a line of effective indoor air quality (IAQ) instruments glorious for his or her straightforward use, consistency and dependability. TSI provides IAQ devices that gauge temperature, wetness (percent RH, wet-bulb, dew point), outdoor air calculations, greenhouse emission, monoxide and mobile particles.

There are two ways of monitoring indoor air in rooms. Hardware based, which is purely related to sensors. Here sensors are deployed in room and which may not result the correct number of occupants in the room and also it's too expensive and the second approach is model-based approach.

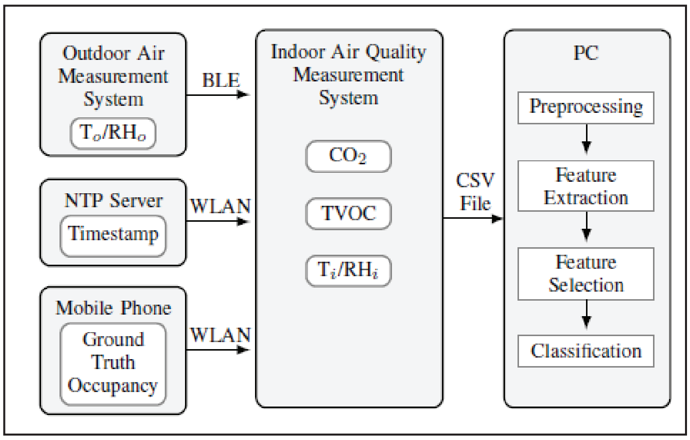


Fig. 1: The Above Diagram Shows the Measurement of Occupancy Detection System [13].

A. Carbon Dioxide

The chemical carbon dioxide gas sensors have the principal advantage of a really low energy consumption and tiny in size. On the negative aspect, it's troublesome to use them in sensible fields because they need a brief life and furthermore as low sturdiness [10]. It's a traditional constituent of the atmosphere at 330-400ppm. SBS symptoms related to CO_2 include headache, fatigue, eye burning and irritation in nose and tract. Materials combined together referred to as aerosol, is an advanced mixture of extraordinarily 2.5 micrometers in diameter and smaller. The scale of the particles is directly coupled to their potential for inflicting health problems. Particles that square measure ten micrometers in diameter or smaller will withstand in throat and nose and enter the lungs. Once in drawn, these particles will have an effect on the center of lungs and cause serious adverse health effects. PM10 can increase the amount of tiny particles and variety of elements, as well as organic chemicals, smoke, soot, dust, salt, acid droplets, metals, and soil. They square measure in two categories of particles i.e. inhalable coarse particles such as those found close to roadways and dust-covered industries that are larger than 2.5 micrometers and smaller or capable of ten micrometers in diameter. It's concentration in indoor air gives an honest indication severity of respiratory illness attacks [2]. Total voltaic compounds:

The summation of all VOCs is known as the total volatile organic compounds (TVOC). The volume of gas per characterization and the aggregate everything being equal (TVOC) is critical impression of the significant natural mixes found in indoor air.
Additionally to serious health considerations, there's the psychological aspect: homes, offices, and different environments that smell clean generally appear additional hospitable than areas with foul odors caused by organic compounds [9].

II. Existing System

Naïve Bayes classifiers are one of the machine learning algorithms where it is a simple probabilistic classifier. It works on Bayes theorem which takes independent assumptions above all the predictors. This model can be built easily and not to complicate because it works on the estimation iterating this helps for large datasets among all the algorithms.
P(A/B) *P(A) =P(B/A)*P(B)
P(A/B) posterior probability
P(B/A) is likelihood which is the probability of predictor given class.
P(B) is predictor prior probability
P(A) class prior probability
In Naïve Bayes classifier there are two models. In Zero model there is no predictor. In OneR model we are trying to find single best predictor [3].

Limitations of Existing System:

- It completely based on assumption of independent variable
- Zero frequency problem

III. Proposed work

In proposed system we use XGBoost machine learning algorithm [14] the main advantage of XGBoost is scalability, so it can learn the data in parallel and distribute computing so it requires less memory.
XGBoost algorithm is also known as ensemble learning method. This algorithm is
being implemented from python package SKLearn.

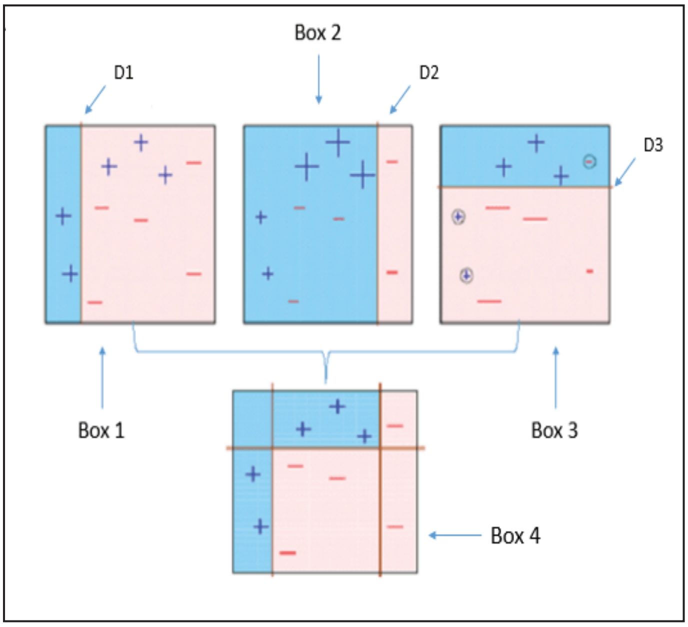


Fig. 2: XGBoost Classification Example

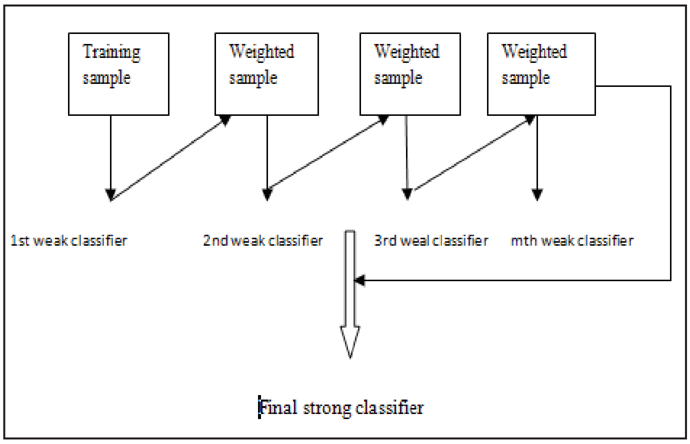


Fig. 3: XGBoost Process Flow

XGBoost uses boosting and bagging technique, Here we first give the total train data to decision tree where it creates the wrongly predicted samples as a first weak classifier, now this first weak classifier is given to the new decision tree it also creates the wrongly predicated samples into another weak classifier this process is done sequentially until the final classifier (strong classifier) is created.
[12] For a given dataset with n instances and with m attributes, the explanatory variables x_i is defined $x_i = (x_{i1}, x_{i2}, \ldots, x_{im})$.
Define the i^{th} objective variable y_i, i=1, 2, 3, …., n
Define the output of tth decision tree:

$$\hat{y}_i^{(t)}$$

The $\epsilon_i^{(1)}$ error between first decision trees output and objective variable yi is following

$$\epsilon_i^{(1)} = \hat{y}_i^{(1)} - y_i$$

Define the k^{th} decision tree: f_k
The predicted value when boosting K times is as

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i)$$

Define the loss function:

$$l(\hat{y}_i, y_i)$$

Our purpose is to minimize the following objective function

$$\mathcal{L}(\phi) = \sum_{i=1}^{I} l(\hat{y}_i, y_i)$$

Define the penalize function:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|\boldsymbol{w}\|^2$$

γ and λ are the hyper parameters,
T is number of tree node,
W is the vector of nodes.
If we add the penalizes function to loss, it helps to smooth the final learnt weight and avoid over-fitting
So, Our new purpose is to minimize the following objective function L (ø):

$$\mathcal{L}(\phi) = \sum_{i=1}^{I} l(\hat{y}_i, y_i) + \sum_{k=1}^{K} \Omega(f_k)$$

XGBoost Features:

Model Features: XGBoost is an implementation of gradient boosting decision tress

There are three forms of gradient boosting:

- Gradient boosting
- Stochastic gradient boosting
- Regularized gradient boosting
- Algorithm Features:
- Handling of missing data values
- Support parallelization of tree construction
- Easy to fit the model on new data because it has continued training

Advantages:

- Execution speed
- Model performance
- Flexibility
- Availability
- Save and reload [5].

Dataset Information:

From [6].we collected the sensor occupancy data where it has total 205354 instances with 32 attributes. In those 32 attributes we have missing values for 10 attributes we eliminate those values. Attributes are named below, outdoor values are completely ignored here.

Table 1: Attribute Information

s.no	Attribute name, units (if any), type
1	wkd {0, 1}, binary
2	Time, seconds, numeric
3	co2,ppm, numeric
4	co2_sma ,ppm,numeric
5	co2_var ,ppm,numeric
6	co2_sma_fd,ppm, numeric
7	co2_sma_sd,ppm, numeric
8	voc, ppm, numeric
9	voc _sma,ppm, numeric
10	voc_var ,ppm, numeric
11	voc_sma_fd ,ppm, numeric
12	voc_sma_sd,ppm, numeric
13	t1, Celsius, numeric
14	t1_sma, Celsius, numeric
15	t1_var,Celsius, numeric
16	t1_sma_fd,Celsius, numeric
17	t1_sma_sd, Celsius, numeric
18	rh1 ,%,numeric
19	rh1_sma,%, numeric
20	rh1_var,%, numeric
21	rh1_sma_fd,%, numeric
22	rh1_sma_sd,%, numeric
23	t12,Celsius, numeric
24	t12_sma ,Celsius, numeric
25	t12_var,Celsius, numeric
26	t12_sma_fd,Celsius, numeric
27	t12_sma_sd,Celsius, numeric
28	rh12 ,%,numeric
29	rh12_sma,%, numeric
30	rh12_var,%, numeric
31	rh12_sma_fd,%, numeric
32	rh12_sma_sd,%, numeric
33	gto {0, 1}

wkd: 1 as occupied and 0 as unoccupied
time_seconds: time of day in second
co2: raw CO_2 measurement value
co2_sma: moving average
co2_var: moving sample variance
co2_sma_fd: first derivative
co2_sma_sd: second derivative
voc: volatile organic compounds
t1: indoor air temperature
rh1: indoor air relative humidity
t2: outdoor air temperature
rh2: outdoor air relative humidity
gto: ground truth occupancy

Table 2: Data Set Information

Dataset/Participant	Vacant	Occupied	Total
T1/A	15503	30281	45784
T2/A	11490	17342	28832
T3/B	9910	10018	19928
T4/B	20495	17033	37528
T5/C	2946	16998	19944
T6/C	14118	5566	19684
T7/D	5359	12275	17634
T8/D	4479	11541	16020
T1-8/A-D	84300	121054	205354

IV. Results

We applied XGBoost algorithm and the results are as follows: From total data 205354 we split into 64% train and 36% test and we have achieved the accuracy of 82.62% and we have improved the accuracy from the existing system where Naïve Bayes has achieved 81.1.

V. Conclusion

In this paper we conclude that the indoor air quality is measured based upon the number of occupants resided in the room. The classification was done through XGBoost algorithm where it learns weak classifiers into strong classifiers.

References

[1] Ravi,"What is a Sensor? Different Types of Sensors, Applications", Electronics Hub, 2017. [Online] Available: https://www.electronicshub.org/different-types-sensors
[2] S. Bhattacharya, S. Sridevi, R. Pitchiah,"Indoor air quality monitoring using wireless sensor network", 2012 Sixth International Conference on Sensing Technology (ICST), 2012.
[3] S.sayad,"Naive Bayesian", Saedsayad.com. [Online] Available: https://www.saedsayad.com/naive_bayesian.htm.
[4] A. Vidhya,"An End-to-End Guide to Understand the Math behind XGBoost", Analytics Vidhya.
[5] D. Flair,"XGBoost in Machine Learning – Features &Importance".

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

[6] [Online] Available: http://bit.ly/occupancy_data

[7] ACEEE,"Indoor Air Quality", ACEEE, 2018.[Online]. Available: https://aceee.org/topics/indoor-air-quality.

[8] TSI,"Indoor Air Quality Meters and Instruments", TSI, Inc. [Online] Available: http://www.tsi.com/indoor-air-quality-meters-and-instruments/.

[9] D. Meyer,"Overview of TVOC and Indoor Air Quality", Zh.idt.com,2018. [Online] Available: https://zh.idt.com/document/whp/overview-tvoc-and-indoor-air-quality.

[10] J. Kwon, G. Ahn, G. Kim, J. Kim, H. Kim,"A study on NDIR-based CO_2 sensor to apply remote air quality monitoring system", 2018.

[11] M. Pathak,"Using XGBoost in Python", DataCamp Community, 2018. [Online] Available: https://www.datacamp.com/community/tutorials/xgboost-in-python

[12] Introduction of Xgboost", Slideshare.net, 2017. [Online] Available: https://www.slideshare.net/palloc/introduction-of-xgboost?qid=95c12b4a-d967-4e19-b4e3-40f4ab8d5ab9&v=&b=&from_search=14.

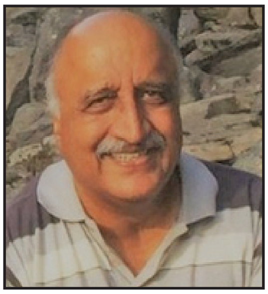
[13] Zimmermann,"Fusion of Non-Intrusive Environmental Sensors for Occupancy Detection in Smart Homes", IEEE, 2017, pp. 10.

[14] T. Chen,"Story and Lessons Behind the Evolution of XGBoost", Homes.cs.washington.edu. [Online] Available: https://homes.cs.washington.edu/~tqchen/2016/03/10/story-and-lessons-behind-the-evolution-of-xgboost.html.

G.Mounika received her Bachelor degree in computer science and engineering from Mahatma Gandhi Institute of Technology College, Hyderabad, India, in 2015, pursuing Masters from Srinidhi institute of science and technology doing research on machine learning.

S.Sai Kumar received his Bachelor degree in computer science and engineering from Mahatma Gandhi Institute of Technology College, Hyderabad, India, in 2016, pursuing Masters from Srinidhi institute of science and technology doing research on machine learning.

Dr.P.K.Sahoo, completed his Ph.D. from Fakir Mohan University, Odisha in Computer Science Engineering. He has 14 years of teaching, research and administrative experience. He has earlier worked as Head of the Department for both CSE and IT in various reputed Engineering Colleges. His Research interest includes Cyber Security, Information Security and Data Mining. He has published around 21 research papers in various reputed journals at national and International levels. He is a certified professional from BalaBit, completed Electronic Contextual Security Intelligence, exam for Intermediate Level (ECSI).



Dr. Kumar joined SNIST from 1999 after leaving BHEL R& D as Additional general Manager. He has Masters and Ph. D, degrees from IIT Kanpur and University of Madras respectively. He now works in the area of Artificial Intelligence involving Neural Networks and Image Processing using pattern recognition methods. He has more than 40 publications in various reputed International Journals and conferences. Many of his publications are referred to by present researchers, even after several decades. He has also won several best teacher awards- the latest being as Best Faculty Award 2013-14, Computer Science from Cognizant for the South India Area.