Indoor Occupancy Detection and Estimation using Machine Learning and Measurements from an IoT LoRa-based Monitoring System

Adeogun, Ramoni Ojekunle; Rodriguez, Ignacio; Razzaghpour, Mohammad; Berardinelli, Gilberto; Christensen, Per Hartmann; Mogensen, Preben

Published in:
Global IoT Summit, GIoTS 2019 - Proceedings

DOI (link to publication from Publisher):
10.1109/GIOTS.2019.8766374

Publication date:
2019

Document Version
Early version, also known as pre-print

Indoor Occupancy Detection and Estimation using Machine Learning and Measurements from an IoT LoRa-based Monitoring System

Ramoni Adeogun*, Ignacio Rodriguez*, Mohammad Razzaghpour*, Gilberto Berardinelli*
Per Hartmann Christensen* and Preben Elgaard Mogensen*†.
*Wireless Communication Networks Section, Department of Electronic Systems, Aalborg University, Denmark
† Nokia Bell Labs, Aalborg, Denmark
E-mail:[ra, irl, mor, gb, phc, pm]@es.aau.dk

***Abstract*—In this paper, we present results on the application of machine learning to the detection of human presence and estimation of the number of occupants in our offices using data from an IoT LoRa-based indoor environment monitoring system at Aalborg University, Denmark. We cast the problem as either binary or multi-class classification and apply a two-layer feed forward neural network to the data. The data used for training, validation and testing of the network comprises of environmental data from the IoT sensors and manual recordings of the door and window states. Results show that the classifier is able to correctly determine occupancy of our offices from the IoT sensor measurements with accuracy up to 94.6% and 91.5% for the binary (presence or absence of persons) and multi-class (no person, one person or two or more persons) problems, respectively. Our analysis also shows that occupancy detection with a network trained either in another room or with single environmental parameter is also possible but with less accuracy.**

***Index Terms*—IoT, machine learning, indoor monitoring, occupancy detection, neural networks, sensor data**

I. Introduction

The Wireless Communication Networks (WCN) Section of the Department of Electronic Systems at Aalborg University, Denmark in collaboration with a number of Danish local industrial partners and the municipality of Aalborg recently deployed an IoT LoRa-based indoor environment monitoring system [1]. An illustration of the system comprising of multi-sensor nodes, LoRa gateways, a back-end server and Grafana open platform based visualization dashboard is shown in Fig. 1.

The sensor nodes measure environmental parameters (including temperature, pressure, Carbon Dioxide (CO_2) and light intensity among others) and transmit same via the internet to the back-end server at AAU, where the data can be extracted and visualized using the dashboard. Following successful deployment of the wireless indoor environment monitoring system at different observation positions in Gigantium - a large sports and culture center in Aalborg municipality, Denmark [1], a number of sensor nodes were installed in office spaces at the FRB building in the Department of Electronic System, Aalborg University in November, 2018. These nodes were used to collect measurement of environmental parameters in our offices over two periods with three weeks duration each, in November and December, 2018.

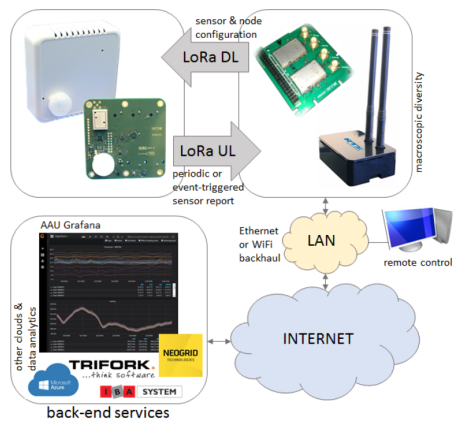


Fig. 1: The LoRa-based indoor monitoring system [1].

Equipped with data from the sensors, our goal is to measure utilization of our office spaces by determining the presence of occupants and estimating the number of persons at a given time. This utilization information can potentially be used to optimize usage of our allocated offices spaces. Moreover, occupancy estimates can also be used for indoor office automation [2]. For example, as input to indoor lighting control system [3], [4] and heat, ventilatilation and air conditioning systems [5], [6]. Environmental parameter measurements contain useful information about occupancy of an enclosed space since human beings affects their environment through, for example, heat generation, Carbon Dioxide (CO_2) emission [7], switching on/off of artificial lighting sources and sound/noise production. While visualization of the sensor data may reveal trends and temporal variations, it is often difficult, if not impossible, to relate these variations to human presence without utilizing data processing tools. Machine learning algorithms can therefore be used to analyze the sensor data and identify patterns. These patterns can then be used to determine occupancy and estimate the number of persons with some degree of accuracy.

In this paper, we investigate the potentials for occupancy detection using data from our environmental measurements. We cast the problem as either a binary or multi-class pattern recognition classification problem and apply a two-layer feed-forward pattern recognition network with sigmoid output neurons [8] to individual parameter measurements and an augmented data. The augmented data is a combination of the automatically collected sensor data (comprising of temperature, pressure, humidity, CO_2, Total Volatile Organic Compounds (TVOC), sound pressure, and PaPIRMotion measurements) and manual recording of the number of persons, window and door position.

The remaining part of this paper is organized as follows. Section II presents a brief overview of the IoT sensor nodes and data collection procedure. Data pre-processing and analysis is presented in Section III. Section IV presents the indoor occupancy prediction procedure using pattern recognition network. Section V presents classification accuracy results and discussion. Conclusions are drawn in Section VI.

II. IoT Sensor Nodes and Data Collection

The indoor monitoring multi-sensor nodes are installed in the WCN offices at the locations indicated in Fig. 2. The datasets used in this paper are those from nodes E2 and FD, which are placed in the section's secretary and a four-persons office, respectively.

A. Sensor Node

Each sensor node consists of nine low-cost sensors: a SENSIRION STS31 high accuracy temperature sensor; a MCP4726 DAC with external custom circuits for sound pressure sensing; a Broadcom APDS-9200 digital UV and ambient light sensor; a BOSCH BME280 combined humidity, pressure, and temperature sensor; an Ams CCS811 ultra-low power digital gas sensor, providing estimated CO2 based on Volatile Organic Compounds (VOCs) measurements; a PaPIRs EKMB1 motion sensor (passive infrared-based), and a ST LSM9DS1TR magnetometer, accelerometer and gyroscope. The sensor nodes capture data and transmit to the back-end server for storage via an IoT LoRa-based network every 5 minutes. Detailed description of the sensor nodes and the LoRa-based IoT network can be found in [1].

B. Data Collection

Sensor data for nodes E2 and FD are extracted from the server and stored into excel files with each row containing the time stamp, seven numerical values (temperature, humidity, pressure, CO_2, TVOC, sound pressure and light intensity) and one categorical value (motion detection). In addition to the extracted sensor data, separate excel files containing three more attributes: number of persons, state of the window (opened (1), tilted (0.5), and closed (0)) and position of the door (opened (1) or closed (0)) were created during the six weeks period when sensor data were collected. Each entry in the manual data has a time stamp corresponding to instants when there is a change in one or more of the attributes.

In Fig. 3, we show variation of the environmental parameters from sensor node E2 aligned with the manually recorded number of occupants over a four days period. Except for pressure, all other parameters show clear variation with the presence of and/or change in the number of persons in the room. Table I gives the statistics - minimum value, mean and maximum value of the data from nodes E2 and FD over whole six weeks duration of the measurement. These values indicate that the likelihood of the sensor data containing outliers is low. The difference in the statistics for data from the two rooms may be associated to the different location and orientation of the rooms as seen in Fig. 2.

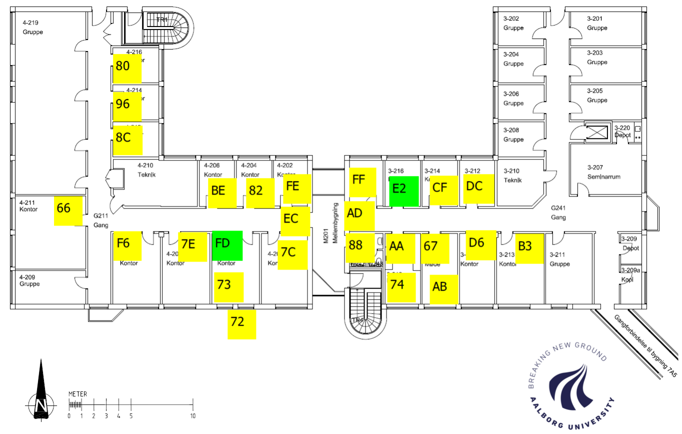
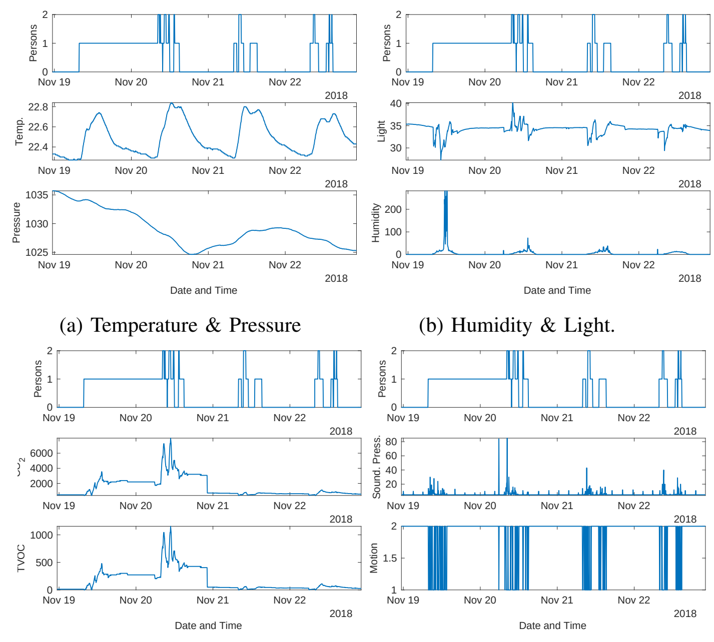


Fig. 2: A portion of the floor plan of AAU Building FRB showing placement of the sensor nodes.



(a) Temperature & Pressure (b) Humidity & Light.

(c) CO_2 & TVOC. (d) Sound Pressure & Motion.

Fig. 3: Variation of environmental parameters from sensor node E2 over a period of about four days.

III. Data Preprocessing

As with any machine learning task, we apply signal processing tools to pre-process the data before applying a pattern

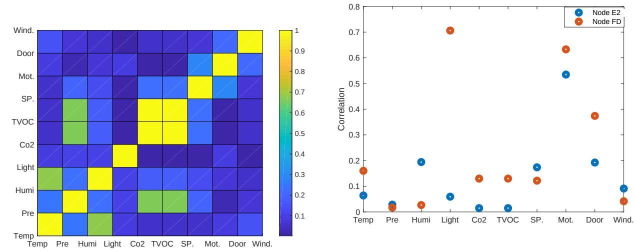
TABLE I: Statistics of IoT sensor environmental data.

	Node E2 (secretary office)			Node FD (4-persons office)		
	Min	Max	Mean	Min	Max	Mean
Temperature (°C)	22.10	24.10	22.70	18.98	23.40	21.80
Pressure (hPa)	1015.70	1040.10	1028.10	977.12	1040.30	1017.70
Humidity (%)	27.30	41.20	35.70	20.03	45.30	29.90
Light (Lux)	0	291.0	18.60	0	274.0	22.50
CO_2(ppm)	400	7832	2009.5	400	7469	1113.2
TVOC (ppb)	0	1132	244.69	0	1076	108.15
Sound Pressure (dB(A)).	4	85	6.6	1	221	6.2

recognition algorithm to determine occupancy and predict the number of persons in the rooms. This step allows us to detect the presence of potential outliers, identify redundancies and generate data for training, validation and testing of the algorithm.

Based on the time stamps, we combined measurements from the sensor with the manually recorded data to form an augmented dataset. To eliminate potential bias, we removed all sensor recording between 6 pm and 6 am from the data. This elimination is done to avoid having too many instances with no occupants. The resulting augmented data contain 3136 and 4091 samples for nodes E2 and FD, respectively. Each sample has eleven input attributes: date and time, temperature, pressure, humidity, sound pressure, light intensity, motion detection, CO_2, TVOC, door and window status and a class attribute (number of persons).

We show similarities between the different parameters in the data in Fig. 4, where we plot the correlation coefficient among input attributes for sensor node E2 in Fig. 4a and the correlation between number of persons in the rooms (target attribute) and input attributes for both sensor nodes in Fig. 4b. We observe in Fig. 4a that CO_2 and TVOC measurements exhibit perfect correlation. This is expected, since the digital gas sensor estimates CO_2 from measurements of Volatile Organic Compounds (VOCs). Either of these attributes can therefore be eliminated from the data without impacting algorithm performance. Some other attributes are also seen to have correlations greater than 0.5. These include temperature and humidity, CO_2 and pressure and TVOC and pressure. The significance of these correlation and possibility for further reduction of the attributes is however, not explored in this work. Fig. 4b shows that all input attributes (measurements) exhibit some correlation with the class attribute (number of people) with light and motion detection measurements having the highest correlation. We further observe that the correlation for some of the attributes differ significantly for the two rooms. For instance, the correlation between number of people and light intensity is approximately 0.7 and 0.05 for sensor nodes FD and E2, respectively. A plausible explanation for this observation as well as the difference in data statistics in Table I is the variation in indoor environment condition in the rooms. These observations raise the question on whether room occupancy in a given room can be performed using machines trained with measurements from a different room.



(a) Correlation between inputs.

(b) Correlation between inputs and target.

Fig. 4: Relationship among inputs and between inputs and target.

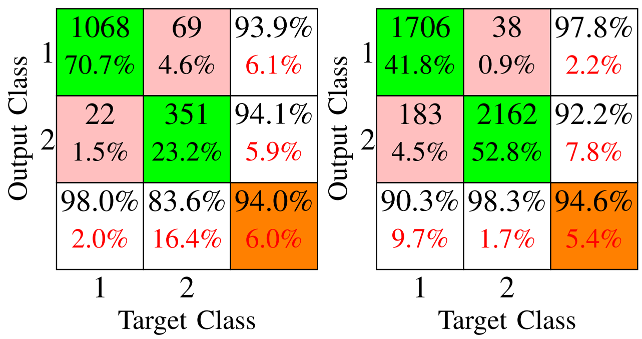
TABLE II: Classification accuracy of applying two-layer feed forward neural network to the data from sensor node FD for both binary and multi-class problems.

	Binary	**Multi-class**
Temperature (°C)	67%	55%
Pressure (hPa)	75.3%	57%
Humidity (%)	71.2%	58.7%
Light Intensity(Lux)	94.3%	73.9%
CO_2(ppm)	63%	52.2%
TVOC (ppb)	63.9%	53.2%
Sound Pressure (dB(A))	53.7%	47.8%
PaPIRMotion	66.1%	58.7%
Door Status	53.8%	50.3%
Window Status	53.7%	46.3%

IV. Room Occupancy Prediction via Machine Learning

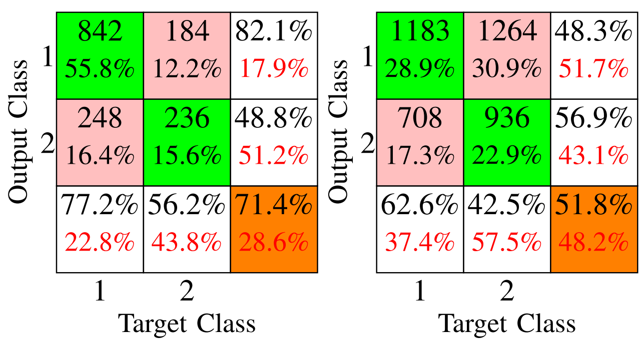
In the machine learning framework, a biologically inspired classification algorithm - a two-layer feedforward neural network (FNN) [8] with sigmoid output neurons, is used to process and learn from the data. The choice of FNN is motivated by the ability of neural networks to learn and model any kind of relationships including non-linear and complex relationships inherent in most real world problems. Except where stated otherwise, network creation, training, validation and testing are performed using MATLAB's neural network pattern recognition tool [9] with default parameters. For each of our experiments, the input data and associated features is randomly grouped into three: 70% for training and 15% each for validation and testing. We evaluate two possibilities for room occupancy detection: a binary problem, which involves detecting the presence or absence of occupants and a multi-class problem, which involves estimation of the actual number of occupants in the rooms.

For the binary problem, we grouped the target attributes (i.e., number of persons, N) into two classes: Class 1 corresponding to instances with $N = 0$ and Class 2 which includes all instances with $N \geq 1$. Since the datasets contain very few instances with 3 or 4 persons, we grouped the data into three classes: Class 1 ($N = 0$), Class 2 ($N = 1$) and Class 3 ($N \geq 2$) for the multi-class problem.



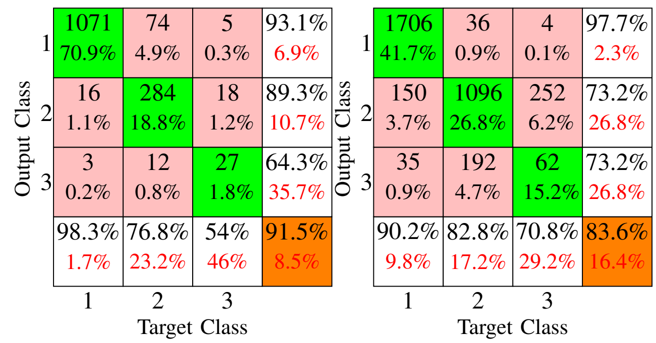
(a) Trained and tested in E2. (b) Trained and tested in FD.

Fig. 5: Intra-room performance of binary room occupancy detection using data from both rooms.



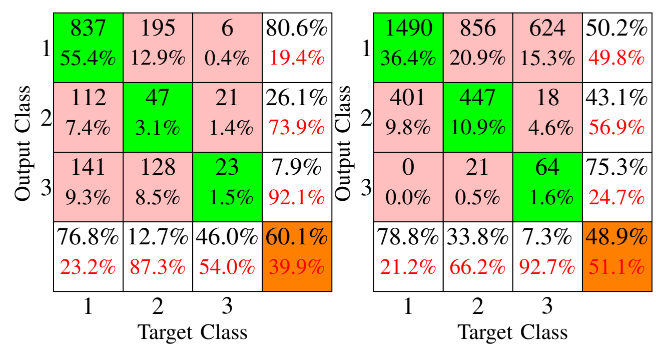
(a) Trained with FD and tested in E2. (b) Trained with E2 and tested in FD.

Fig. 6: Inter-room performance of binary room occupancy detection using data from both rooms.



(a) Trained and tested in E2. (b) Trained and tested in FD.

Fig. 7: Intra-room performance of multi-class room occupancy estimation using data from both rooms.



(a) Trained with FD and tested in E2. (b) Trained with E2 and tested in FD.

Fig. 8: Inter-room performance of multi-class room occupancy estimation using data from both rooms.

V. Results and Discussion

We evaluate performance of the network using accuracy and miss rate metrics. Accuracy and miss rate show the percentage of entries that are correctly and incorrectly classified, respectively. Table II presents the accuracy of both binary and multi-class occupancy prediction with single input attributes using data from sensor node FD. It shows that accuracy of the network varies with each of the attributes (environmental parameters and door/window status) as input. For binary(multi-class) problem, light intensity and window status yield the highest, 94.3% (73.9%) and lowest, 53.7% (46.3%) accuracy, respectively. Table II also shows that all input attributes yield higher accuracy for the binary problem. We will now show network performance results with all attributes except TVOC measurements as input. TVOC measurements are eliminated considering the linear relationship with CO_2 measurements.

Fig. 5 presents the overall confusion matrix for the binary problem with all input attributes. The rows correspond to the output (i.e., estimated) class and the columns to the target (i.e., true) class. The diagonal and off-diagonal cells indicate the observations that are correctly and incorrectly classified, respectively. Each cell contains both the number of observations and percentage of the total number of observations. The column on the far right shows the percentages of all the observations predicted to belong to each class that are correctly and incorrectly predicted. The row at the bottom shows the percentages of all observations in each class that are correctly and incorrectly classified. The bottom right cell (highlighted orange) gives the overall classification accuracy. Fig. 5 shows that the network prediction for the binary problem is very accurate with accuracy of 94% and 94.6% for sensor nodes E2 (secretary office) and FD (four-persons office), respectively.

In Fig. 6, we plot the confusion matrix for the binary problem with training and test samples from either of the two rooms. Fig. 6a (6b) shows the confusion matrix with data from node FD (E2) and E2 (FD) used for training and testing, respectively. Compared to Fig. 5, predictions from the network is less accurate with overall accuracy of 71.4% (51.8%) for a network trained with FD (E2) dataset and tested using observations from node E2 (FD). This appears reasonable considering the differences in statistics of observations from the two rooms in Table I and the correlation in Fig. 4b.

We show the performance of the network for multi-class classification in Fig. 7b, where we plot the confusion matrix obtained with data from both sensor nodes. The classification accuracy is 91.5% and 83.6% for node E2 and FD, respectively. While this accuracy may be reasonable considering the limited amount of observations, the observed accuracy is slightly lower than that of the binary problem in Fig. 5. Fig. 7b also show that classification accuracy differs significantly for each of the three classes. For example, while class $1(N = 0)$ has accuracy of 98.3% (90.2%), class 3 ($N \geq 2$) has much lower accuracy of $54\%(70.8\%)$ for node E2 (FD). A plausible explanation for this performance variation is the proportion of each class contained in the dataset.

Finally, we plot the confusion matrix for multi-class with the network trained and tested using data from different rooms. Compared to Fig. 7b where training and testing is done using data from the same room, the performance is much worse with accuracy of 60.1% (48.9%) for a network trained with FD (E2) and tested using E2 (FD) dataset.

VI. Summary and Conclusion

We have presented results on application of machine learning to indoor occupancy estimation using sensor data from a wireless IoT indoor monitoring system deployed at Aalborg University, Denmark. We combined multi-sensor measurements of eight parameters with manual recording of the number of persons, window position and door status to form an augmented data set. We cast the problem as either a binary or multi-class pattern recognition classification problem and applied a two-layer feed-forward pattern recognition network with sigmoid output neurons.

With both training and testing data from the same room, the network is able to correctly determine occupancy of selected offices from the IoT sensor measurements with accuracy up to 94.6% and 91.5% for the binary and multi-class problems, respectively. Occupancy detection with a network trained either in another room or with single environmental parameter is also possible but with less accuracy. Further improvement in classification performance via for example, network optimization, larger training datasets and usage of other algorithms will be considered in our future work.

References

[1] I. Rodriguez, M. Lauridsen, G. Vasluianu, A. N. Poulsen, and P. Mogensen, “The Gigantium Smart City Living Lab: A Multi-Arena LoRa-based Testbed,” in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2018, pp. 1–6.

[2] Z. Han, R. X. Gao, and Z. Fan, “Occupancy and indoor environment quality sensing for smart buildings,” in *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, May 2012, pp. 882–887.

[3] M. T. Mulia, S. H. Supangkat, and N. Hariyanto, “A review on building occupancy estimation methods,” in *2017 International Conference on ICT For Smart Society (ICISS)*, Sep. 2017, pp. 1–7.

[4] “Energy intelligent buildings based on user activity: A survey,” *Energy and Buildings*, vol. 56, pp. 244 – 257, 2013.

[5] J. Ahmad, H. Larijani, R. Emmanuel, M. Mannion, and A. Javed, “Occupancy detection in non-residential buildings a survey and novel privacy preserved occupancy monitoring solution,” *Applied Computing and Informatics*, 2018.

[6] M. Jin, N. Bekiaris-Liberis, K. Weekly, C. J. Spanos, and A. M. Bayen, “Occupancy detection via environmental sensing,” *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 2, pp. 443–455, April 2018.

[7] B. Abade, D. Perez Abreu, and M. Curado, “A Non-Intrusive Approach for Indoor Occupancy Detection in Smart Environments,” *Sensors*, vol. 18, no. 11, 2018.

[8] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85 – 117, 2015.

[9] MATLAB, “Classify Patterns with a Shallow Neural Network,” https://se.mathworks.com/help/deeplearning/gs/classify-patterns-with-a-neural-network.html#f9-48630.