Building Occupancy Estimation using Supervised Learning Techniques

Claudia Chiţu*†, Grigore Stamatescu*‡, Alberto Cerpa†
*Department of Automatic Control and Industrial Informatics
University Politehnica of Bucharest, Romania
†Department of Electrical Engineering and Computer Science, University of California Merced, USA
‡Institute of Technical Informatics, Graz University of Technology, Austria

***Abstract*—Smart buildings viewed as cyber-physical systems are currently a growing research topic oriented towards collaborative groups of buildings. Since buildings consume significant amount of energy, research efforts have concentrated to make them more efficient, in particular the Heating, Ventilation and Air-Conditioning (HVAC) systems that represent more than 40% of the buildings' energy budget. A key piece of information that facilitates the design of energy efficient HVAC systems, in particular in commercial buildings, is the knowledge of the real-time and predicted occupancy, which would allow an automatic control process to balance the trade-off between energy use and quality of comfort. In practice however, occupancy counting devices are not being wide-spread deployed in the market, so in order to move forward, we believe it is important to estimate occupancy using *existing* sensors currently deployed in buildings. In this work, we propose to use a combination of sensor data currently available in buildings, such as CO_2 data and airflow, and develop a supervised learning framework that uses existing data to estimate occupancy. We developed two data-driven techniques based on Random Forest (RF) and KNN algorithms to estimate occupancy based on data collected from 4 rooms. Our results show an average RMSE occupancy error that varies from 3.10 to 11.21 for RF (depending on the room) and 2.96 to 8.46 for KNN, with best case results of 1.08 and 0.97 respectively. We believe that our framework can be integrated into existing Building Management Systems (BMS) control processes to improve energy efficiency in smart buildings.**

***Index Terms* - Smart Buildings, Random Forest, KNN, occupancy estimation**

I. Introduction

Precise occupancy estimation in buildings has many applications, enabling smart Heating, Ventilation and Air Conditioning (HVAC) system utilization [1], [2] as well as dynamic and real-time thermal comfort [3], [4], [5] and better building control [6]. Commercial buildings have great potential to run control processes based on real-time and predictive occupancy, improving the quality of service perceived by users in terms of temperature comfort and healthy air ventilation, while minimizing energy use. In commercial buildings, HVAC systems consume around 42% of the overall energy [7], and the Building Management Systems (BMS) cannot efficiently reduce this consumption with modern control techniques due to lack of information with regards to two main aspects: lack of knowledge regarding the actual location of the users inside the buildings at any time (occupancy patterns), and lack of knowledge regarding their subjective demands (comfort levels). In this work, we concentrate on data-driven techniques to address the occupancy modelling challenges when there is no infrastructure specifically deployed to provide reliable direct measurements, and try to leverage *existing* data sources, such as CO_2 data and airflow to address the problem.

Occupancy estimation is a challenging issue due to multiple factors: i) there is currently no wide-spread adoption of special sensors in buildings to directly measure occupancy in each zone; ii) ground truth occupancy data, critical for evaluation purposes, is obtained using video cameras, which is still difficult to process and violates data and privacy concerns of the building occupants; iii) the algorithms and techniques used for occupancy estimation need to deal with poor data quality from available sources, which exacerbates the problem of feature selection and minimization of error for occupancy.

In our work, we present a data-driven framework based on statistical learning techniques that addresses some of the shortcomings mentioned above. We evaluate two techniques to estimate the real-time and predictive occupancy based on existing data sources in the building: (a) Random Forests (RF) and (b) K-Nearest Neighbors (K-NN). Random Forests have not been extensively exploited in occupancy estimation and prediction, but it is a known and proved technique with high performance that corrects for the tendency of baseline decision trees for overfitting the training set. K-NN is a non-parametric method that is straight forward to implement. It is however sensitive to the local structure of data and it is suitable to be implemented on large datasets with few attributes. The only possible drawback of computational cost does not apply to our case since we consider a sliding window method for data training and testing. Furthermore, both methods can run in any modern desktop machine without the need of cloud services, a situation often encountered in small and medium sized building without extensive IT infrastructure.

Our work builds upon the latest research results in the topic of smart buildings and occupancy estimation, highlighting the impact of quality of the data and the importance of proper occupancy behavior modeling. This paper provides the following contributions:

- We developed, tested, and evaluated two machine learning approaches for occupancy prediction using data from an office building [8] with regards to CO_2 concentration and ventilation airflow, as an alternative to situations when dedicated occupancy sensors are

not in place.
- We assessed the validity of our results for occupancy profile identification using a sliding window splitting method for data feed from several rooms in University of Southern Denmark as a use case [8].

Our paper is organized as follows: in Section II, we present related work which defines the current scientific context of our contributions. Section III describes the data driven approaches used for occupancy estimation using a sliding window methods for the input data stream. In Section IV we present the results of our evaluation, discussing the insights provided by our solution. We discuss significant insights, emphasizing the benefits of using our proposed solution. Finally, in Section V we conclude, summarizing the results and describing areas for future research directions.

II. Related Work

Novel smart building technologies affect multiple stakeholders, from building facility teams and building owners, to energy suppliers, research engineers, and occupants of modern buildings. There is a trend to start using sensors to detect human presence and integrate them with the local information systems in order to cut operational costs.

In spite of these advancements, the energy sector still faces a performance gap in the building domain related to multiple factors, in particular design factors, construction factors and operation modes. Energy efficient operation of buildings could take advantage of recent advancements in the field. Recent studies [9], [10] show that the energy performance gap in buildings is sustained mainly by lack of occupancy monitoring, occupants behavior models and inappropriate building management and control strategies. Furthermore, in the literature it is emphasized the positive impact on energy performance that energy usage monitoring has within a dynamic schedule modeling [11].

Applying schedules based on estimated occupancy show promising energy savings [12] illustrating ways of implementing them such that temperature and airflow setpoints are set to vary on a larger range. This is explained by the fact that persons induce loads with their dissipated body heat and appliances used, which trigger the HVAC system to send proper commands towards the variable air volume (VAV) units that deliver the cooling/heating energy to the thermal zones.

Occupancy estimation and prediction [13] are often used to optimize building systems employing Model Predictive Control [14], [15]. The data sources for the human presence detection are from new wireless sensor networks [16] such as CO_2, PIR, infrared sensors but also from existing infrastructure: smart meters, WiFi, and HVAC sensors.

Modeling methodologies from recent studies conducted to asses the room usage percentage, report a combination of multiple data sources types and broad palette of algorithms. For instance, in [17], the authors used an infrared sensor grid to detect human presence which feeds a Markov model chain algorithm. To reveal patterns in human behavior within office buildings, some researchers from Singapore [18] proposed fusion framework using particle filter algorithm with improvements of 5-14% for estimation accuracy among several methodologies. Another group of researchers [19], from China, propose a solution consisting in the combination of K-NN with K-means clustering to detect high resolution occupancy level using WiFi and Bluetooth Low Energy technologies. They report thermal load savings up to 14.16% compared with actual occupancy at 50% humidity and 25°C temperature. From the same family with K-NN models, the Random Forest (RF) models were applied and returned promising results for occupancy tasks [20], [21].

The state-of-the-art solutions previously presented showed different sensing technologies and data-driven techniques to estimate occupancy in commercial buildings. We note that sometimes, algorithms for occupancy estimation are difficult to be adopted and transferred to other spaces/buildings due to the lack of similar sensor networks or engineering features. Thus, we employ supervised learning techniques using existing data sources which are present in most of university buildings.

III. Occupancy Models

In this section, we explain the algorithms and techniques used to solve the occupancy estimation problem. In our work, we use RF and K-NN classification algorithms to address the estimation of the occupancy levels of the rooms/zones based on the several inputs that are usually available in commercial buildings. We use as inputs the value of the CO_2 concentration, as measured by sensors, and ventilation airflow, as measured by the position of the airflow damper, in each room/zone. The output of our classifiers is the total number of occupants in each room/zone, which is an integer number. We used the open dataset provided by researchers from the University of Southern Denmark [8] to train and test our classifiers.

The use of CO_2 sensors as input was motivated by the availability of these type of sensors in many commercial building installations. Although CO_2 sensors have shown to have several drawbacks, like long response time and the dependency of location based on the space layout, lately, with proper calibration, their accuracy and precision have been increased in practice.

From our data set, we use data from 4 rooms from a university building with approximately 1000 occupants on normal workdays, data collected in the spring of 2017 for 15 days, with a temporal resolution of one minute. In total, the data set has 21,600 readings per room for each sensor, which means 259,200 records of 3 sensor types: CO_2 sensor that measures the levels concentration in ppm, damper openness to represent the ventilation airflow, coming from control system devices, and occupant counts. Room identifiers have been removed due to the anonymity requirements.

The modeling process starts with data profiling and analysis, which is usually a time demanding phase. Although the authors of [8] specified that data has been cleaned, we performed a data summary and a visual

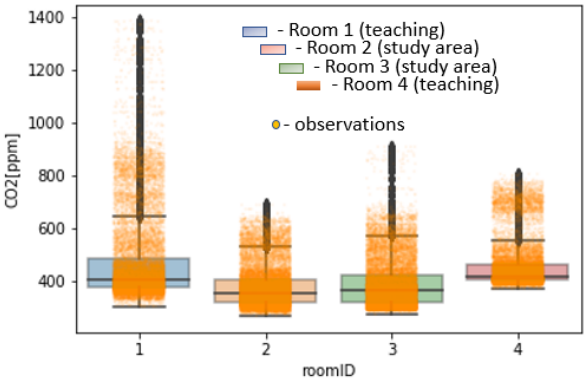

Fig. 1. CO_2 data analysis

inspection tasks. Figure 1 shows a boxplot diagram for CO_2 concentration levels for the 4 rooms. This type of plot is useful to asses some key characteristics of our dataset: outliers, median values, data skewness.

Despite the fact that data does not have a normal distribution and many outliers could be seen in this plot, we checked the nature of these odd points and found them valid points to be considered. According to the Labour Inspectorate in Denmark a guiding index for values of CO_2 levels in classroom is recommended to be 1200 ppm [22] and the maximum value for CO_2 in the plot does not exceed 1390 ppm, though a general drowsiness might be felt at this level. Furthermore, we found that the values greater than 1000 ppm represent only 0.96% from all the records, 1000 being compliant even with the standards developed by the American Society Of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) [23] and the Occupational Safety and Health Administration (OSHA). In addition, adverse effects may be expected at concentration values bigger than 2500 ppm but lower than 5000 ppm. The values for CO_2 concentration in classrooms should be checked for every country since its upper limit of acceptance, not debating the comfort, is regulated differently in different countries. This was the only concerning parameter, because the damper position, which is our second parameter/input, is mechanically limited and controlled by the BMS. This visualization points out that Room 2 and Room 3, which are the same size present similar behavior, in contrast with Room 1 and Room 4. This difference can be explained in the last column of Table I, because the number of occupants in Room 1 is larger with 58% comparing with Room 4, at peak usage.

The occupant counts are obtained from processed data captured with PC2 3D stereo vision cameras and then transformed to respect the individual identity and European regulations, with the aid of Probabilistic Fusion Algorithm; the results are considered ground truth due to 0.075 RMSE accuracy obtained.

Two of the rooms are teaching rooms and the other two are study zones. Their specifications are presented in Table I, where the actual maximum occupancy represents the maximum number of persons observed in the room, selected from the entire period of data collection.

TABLE I
SPACE DETAILS FOR DATA COLLECTION

RoomID	Room type	Size[m^2]	Capacity[seats]	Actual max. occupancy
1	teaching	139	84	67
2	study area	125	32	28
3	study area	125	32	35
4	teaching	139	84	39

The RF model is based on the bagging approach, which means an average model for random samples; this way variance reduction and over-fitting are avoided. Having a training dataset as input $X = \{x_1, x_2, ..., x_n\}$ and a testing dataset $Y = \{y_1, y_2, ..., y_n\}$ the RF model assumes a training of a classification tree T_b on a sample with replacement dataset from X_b and Y_b with $b = \{1, ..., B\}$ where B is the number of the bagging repetitions. The output of the algorithm is the ensemble of trees $\{T_b\}_1^B$. In the case of the classification task, the prediction for a new unseen sample, x, is:

$$\hat{T}_{rf}^B(x) = MajorityVote\{\hat{T}_b(x)\}_1^B \tag{1}$$

For the RF model, the Gini index (G) is used, as a less computational complex metric, as criterion to decide splits in dataset. This leads to finding the probabilities for each class within the classification process:

$$G = 1 - \Sigma_{i=1}^n (p_i)^2 \tag{2}$$

Visualization of how decision trees are working in the RF, help to understand how the target variable is predicted, in our situation, the number of occupants. We illustrate a small tree in Figure 2 for data from the third room.

The RF model is built upon votes from the decision trees from which it is made. For the root node, CO_2 is the variable to split the node on with 10,912 data points, also called samples. On the second row of nodes, the 'Damper' variable is used to make the split again. Because we illustrate this example for Room3, we have 35 maximum occupants, so we have 35 values in the 'value' vector, which is in conformity with the specifications from Table I. Looking at the leaves row (the bottom line) in Figure 2, on the left side, the first value from this vector is 7977, the largest one in the array, which means class '1' (occupant). Looking at the right side leaf, the maximum value in the array named 'value' is 139, which corresponds to class '13' (occupants count).

In the case of RF, we investigated the impact that the feature selection we chose has on the algorithm performance and its influence is presented in Table II. The feature importance in Table II is for the case when the algorithm ran with the entire dataset, split in two, with 20% data used for testing. Besides this, the number of trees is set to 50, but we tested different values of different orders of magnitude. Our results show that both features have great importance for a good overall result, and when

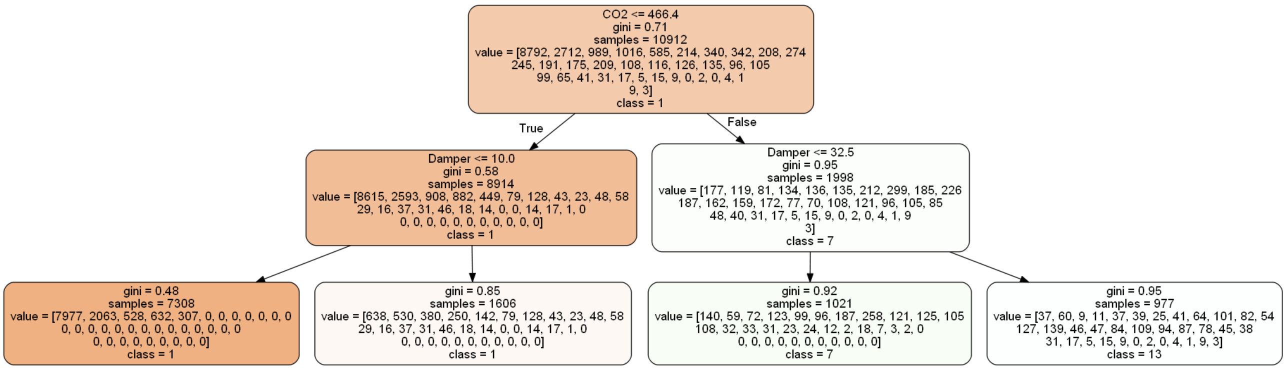

Fig. 2. Decision tree in RF model

TABLE II
FEATURE SELECTION IMPORTANCE FOR RF

RoomID	CO_2 importance	Damper openness importance
1	0.4823	0.5176
2	0.6188	0.3811
3	0.5344	0.4655
4	0.4929	0.5070

the room size is smaller (see Table I), CO_2 concentration has a slightly bigger influence than the damper openness feature.

The K-NN classifier is based on a similar method to the RF; k training points closest in distance to a given point are used to get the majority vote among the k neighbors.

For the K-NN parameter tuning, we selected an array of values for k and ran the algorithm along the range for different rooms with different training and testing datasets. The values of k are different for each room, and can be significantly different in magnitude (e.g. 500 vs 2,500), being dependent on the occupant behavior, as shown in Figure 3. The k parameter is dependent on the room usage, a pattern that is reflected on the value range that works for Room3 but does not necessarily fit the tuning for Room1. Thus, the tuning parameter step is not a straight forward step since the k value should be tested intensively to find how it impacts the accuracy of predictions.

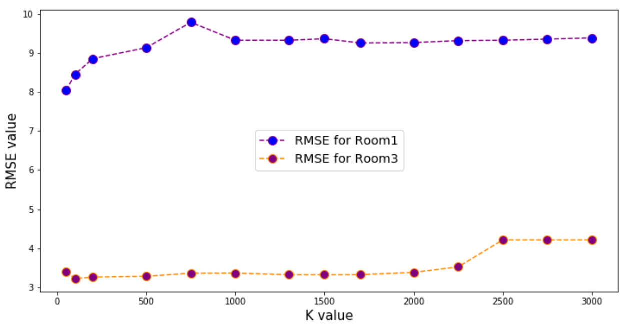

Fig. 3. Parameter selection for k-NN

The k-NN algorithm was implemented using the Euclidian distance, where p and q are two sets of data points:

$$d(p, q) = \sqrt{\Sigma_{i=1}^{n}(q_i - p_i)^2} \tag{3}$$

We implemented the algorithms in Python using Pandas and Scikit-learn libraries. The full dataset that we processed is about 1MB, and the Pandas software framework is suitable for our case, since it is very efficient with time series data. For the RF algorithm, with 50 trees, considering the case of the 4 datasets obtained with window splitting method, the training time ranges from 0.11 seconds to 0.14 seconds for the 4 rooms; the same dataset, with K set to 1000, the K-NN algorithm performed the training phase in 0.02 seconds up to 0.07 seconds. Due to minimum resources employed and an open dataset, the entire use case presented in this paper can be fully replicated.

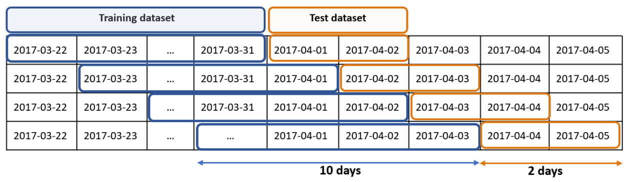

Fig. 4. Data split using sliding window

IV. RESULTS AND DISCUSSION

In this section, we show the findings of our algorithms after tuning RF and K-NN models, comparing their behaviors for several datasets across the 4 rooms, and discuss the limitations of our study.

In our evaluation we worked on the custom data split using a sliding window method similar to the one from [13] to mimic natural behavior of temporal occupancy dependence, considering 10 days for training phase for both algorithms and 2 days for testing. Using this approach, we obtained 4 datasets as shown in Figure 4.

In our evaluation, we use the following RMSE formula:

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_{actual} - y_{pred})^2} \tag{4}$$

Figure 5 shows the occupancy ground truth and the occupancy estimation predicted by both the RF and k-NN model for one hour at 1 minute sample granularity. The models were trained/tested with a split of 20% for testing, with 50 trees for RF and k = 1000 neighbors for k-NN.

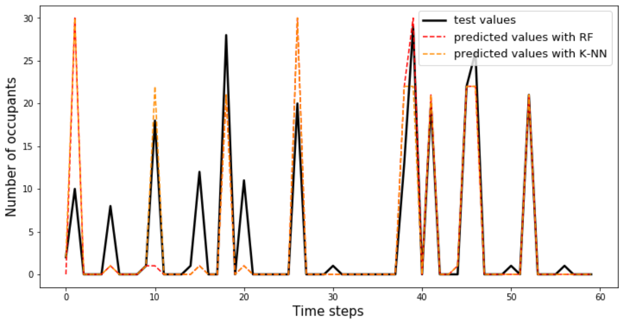

Fig. 5. Occupancy detection for Room4 using RF versus K-NN

TABLE III
RMSE AVERAGE VALUE FOR RF AND K-NN MODELS

RoomID	RF RMSE	K-NN RMSE
1	11.21	8.46
2	3.39	3.15
3	2.34	2.08
4	3.10	2.96

The second model achieved a slightly better performance: 4.03, whilst the first one only 4.13 in terms of RMSE.

Figure 6 illustrates the behavior of RF versus k-NN models using the sliding method with the first 10 days as training window and the next 2 days as testing window. The number of trees remains set to 50 for RF, but the parameter k is set to 2500. For the first half of the plot, the horizontal line means that no occupants were in the Room 1 for the first test day, which is a teaching room; this could be explained with a weekend day. In this case, again, K-NN returned better results than RF. With k = 2500, the RMSE value is 4.52 in this case and 5.1 for k = 2000, although for the case without sliding window split, a smaller value for k indicated that RMSE value would drop (see Figure 3).

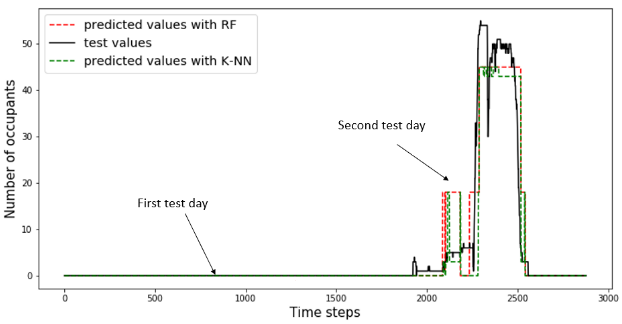

Fig. 6. Occupancy count for Room1 using RF versus K-NN with sliding window

Figure 7 shows the RMSE errors for the two models in each of the 4 rooms using the datasets with the sliding window split as shown in Figure 4, we presented the evaluation with RMSE in Figure 7. We clearly see that the RMSE tend to remain relatively small for rooms 2-4, but tend to increase significantly for room 1. This is the result of the sliding window split, which may not include the cases of maximum occupancy in the training set when doing the split.

Table III shows the average RMSE error for both RF and K-NN for the different rooms. Overall, the K-NN technique performs slightly better in all cases. When re-normalizing the average RMSE error based on maximum occupancy for each room (see Table I), we see that the occupancy error varies from 5% to 16% depending on each room. A basic observation of our results is that the larger the actual maximum occupancy, the larger the expected occupancy error, as it can be seen by the disparity of the average RMSE results between room 1 (much higher maximum occupancy) vs. the rest of the rooms.

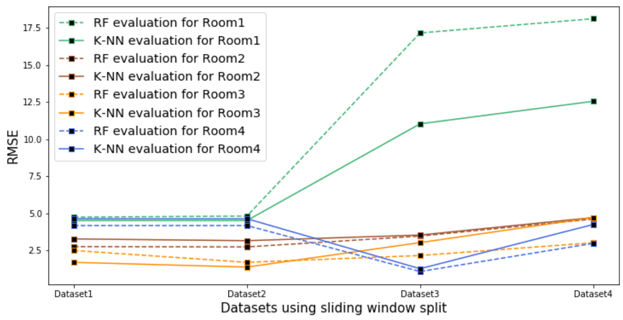

Fig. 7. Sliding window dataset split evaluation benchmark for the 4 rooms

Although room temperature may increase the accuracy of occupancy estimation, even without it, our results still show that conditioning a room including schedules based on estimated occupancy using K-NN models could potentially reduce the overall energy consumption.

We believe that energy savings can be achieved in part by not ventilating the room for maximum capacity, but balancing the k-NN model occupancy output with a safety guard band that is proportionally to the worst case RMSE. For instance, if k-NN shows an occupancy error of 10 persons over the ground truth in the worst scenario, we could add a safety margin proportionally to the RMSE in order to ensure comfort limits are still satisfy in any case, but still remaining below the maximum room capacity and producing energy savings. This is suitable for lecture rooms or large office spaces where occupants count is a large number, so estimation could float on a bounding range around the actual occupancy without a dramatic impact. Although this approach is suitable for lecture rooms or other large similar spaces, testing of these techniques to smaller spaces should be properly evaluated.

There are several limitations of our results that we would like to point out. First, the data used comes from only one season, spring, and a single city, hence we would need further data sources to check if these techniques generalize well under different conditions. Second, our study could be improved with a section dedicated to the accuracy of the CO_2 concentration levels, which are dependent on the sensor calibration and placement of the sensor. Finally, considering the applicability of the temperature and ventilation HVAC control, it is important to distinguish between large zones/rooms occupied by tens of users and small offices that tend to be occupied by a very small number of users. This is important, because in the zero-one occupancy boundary, there is a discontinuity point in the design. The optimal decision for an empty room is to let it float temperature wise in order to save energy, but if the room has a single user, the BMS must do temperature conditioning to satisfy user comfort requirements. This means that the occupancy error in our models would be too large to address this condition. We believe that additional sensor types, like PIR sensors commonly used for lighting control, could complement the occupancy models we developed here in these situations.

V. Conclusions and Future Work

In this paper we have presented and evaluated two supervised learning techniques to perform occupancy estimation in commercial buildings based on readily available existing sensors and data sources. By using data from CO_2 sensors and damper position for ventilation airflow as input, and training both Random Forest (RF) and K-Nearest Neighbor (K-NN) classifiers, we illustrate that average RMSE occupancy errors between 3 and 11 occupants are possible, leading to occupancy errors between 5% to 16% of the maximum occupancy. We assessed the validity of our results with a sliding window splitting method that generalize well with the data for different zones/buildings. In the future, we plan to integrate these techniques with Model Predictive Control (MPC) schemes, that would allow finer control of zone temperature and ventilation levels based on estimated occupancy. This work is the initial foray into data-driven modeling techniques for building occupancy using existing building sensor and data, that would pave the way for more efficient modalities to perform HVAC control in commercial buildings and balance the key trade-offs between energy savings and user comfort.

References

[1] V. L. Erickson, M. Á. Carreira-Perpiñán, and A. E. Cerpa, "Observe: Occupancy-based system for efficient reduction of hvac energy," in *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*. IEEE, 2011, pp. 258–269.

[2] V. L. Erickson, S. Achleitner, and A. E. Cerpa, "Poem: Power-efficient occupancy-based energy management system," in *Proceedings of the 12th international conference on Information processing in sensor networks*. ACM, 2013, pp. 203–216.

[3] V. L. Erickson and A. E. Cerpa, "Thermovote: participatory sensing for efficient building hvac conditioning," in *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. ACM, 2012, pp. 9–16.

[4] G. Stamatescu and A. Cerpa, "Zone-level agreement by consensus for building thermal energy management," in *Control and Automation (ICCA), 2016 12th IEEE International Conference on*. IEEE, 2016, pp. 377–382.

[5] D. A. Winkler, A. Beltran, N. P. Esfahani, P. P. Maglio, and A. E. Cerpa, "Forces: feedback and control for occupants to refine comfort and energy savings," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 1188–1199.

[6] A. Beltran, V. L. Erickson, and A. E. Cerpa, "Thermosense: Occupancy thermal based sensing for hvac control," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013, pp. 1–8.

[7] U.S. and D. of Energy, "2011 buildings energy data book," 2012.

[8] K. Arendt, A. Johansen, B. N. Jørgensen, M. B. Kjærgaard, C. G. Mattera, F. C. Sangogboye, J. H. Schwee, and C. T. Veje, "Room-level occupant counts, airflow and co2 data from an office building," in *Proceedings of the First Workshop on Data Acquisition To Analysis*, ser. DATA '18. New York, NY, USA: ACM, 2018, pp. 13–14. [Online]. Available: http://doi.acm.org/10.1145/3277868.3277875

[9] M. Jradi, K. Arendt, F. Sangogboye, C. Mattera, E. Markoska, M. Kjærgaard, C. Veje, and B. Jørgensen, "Obepme: An online building energy performance monitoring and evaluation tool to reduce energy performance gaps," *Energy and Buildings*, vol. 166, pp. 196–209, 2018.

[10] M. Schmidt and C. Åhlund, "Smart buildings as cyber-physical systems: Data-driven predictive control strategies for energy efficiency," *Renewable and Sustainable Energy Reviews*, vol. 90, pp. 742–756, 2018.

[11] B. Frei, C. Sagerschnig, and D. Gyalistras, "Performance gaps in swiss buildings: an analysis of conflicting objectives and mitigation strategies," *Energy Procedia*, vol. 122, pp. 421–426, 2017.

[12] O. Ardakanian, A. Bhattacharya, and D. Culler, "Non-intrusive occupancy monitoring for energy conservation in commercial buildings," *Energy and Buildings*, vol. 179, pp. 311–323, 2018.

[13] F. C. Sangogboye, K. Arendt, A. Singh, C. T. Veje, M. B. Kjærgaard, and B. N. Jørgensen, "Performance comparison of occupancy count estimation and prediction with common versus dedicated sensors for building model predictive control," in *Building Simulation*, vol. 10, no. 6. Springer, 2017, pp. 829–843.

[14] G. Neculoiu, G. Stamatescu, and V. Sgarciu, "Thermal control strategies applied in buildings with intermittent heating," *UNIVERSITY POLITEHNICA OF BUCHAREST SCIENTIFIC BULLETIN SERIES C-ELECTRICAL ENGINEERING AND COMPUTER SCIENCE*, vol. 78, no. 4, pp. 31–42, 2016.

[15] G. Neculoiu, V. Dache, G. Stamatescu, and V. Sgarciu, "Model predictive control applied for building thermal control," in *Electrical Machines & Power Electronics (ACEMP), 2015 Intl Conference on Optimization of Electrical & Electronic Equipment (OPTIM) & 2015 Intl Symposium on Advanced Electromechanical Motion Systems (ELECTROMOTION), 2015 Intl Aegean Conference on*. IEEE, 2015, pp. 363–368.

[16] G. Stamatescu, C. Chiţu, C. Vasile, I. Stamatescu, D. Popescu, and V. Sgârciu, "Analytical and experimental sensor node energy modeling in ambient monitoring," in *Industrial Electronics and Applications (ICIEA), 2014 IEEE 9th Conference on*. IEEE, 2014, pp. 1615–1620.

[17] C. Chitu, M. I. Khadem, and V. Sgarciu, "Predictive modeling of occupancy patterns in smart buildings," in *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2017 9th IEEE International Conference on*, vol. 2. IEEE, 2017, pp. 622–627.

[18] Z. Chen, M. K. Masood, and Y. C. Soh, "A fusion framework for occupancy estimation in office buildings based on environmental sensor data," *Energy and Buildings*, vol. 133, pp. 790–798, 2016.

[19] W. Wang, X. Xu, H.-H. Wei, B. Ren, and J. Chen, "Modeling occupancy distribution in large building spaces for hvac energy efficiency," *Energy Procedia*, vol. 152, pp. 1230–1235, 2018.

[20] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and co2 measurements using statistical learning models," *Energy and Buildings*, vol. 112, pp. 28–39, 2016.

[21] H. Zou, Y. Zhou, J. Yang, W. Gu, L. Xie, and C. Spanos, "Freedetector: Device-free occupancy detection with commodity wifi," in *Sensing, Communication and Networking (SECON Workshops), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–5.

[22] Tinytag. (2018) Monitoring carbon dioxide levels to help maintain air quality and assess hvac performance. [Online]. Available: https://www.geminidataloggers.com/info/carbon-dioxide-air-quality-data-logger

[23] E. ASHRAE, "Ashrae/iesna 90.1 standard-2007-energy standard for buildings except low-rise residential buildings," *American Society of Heating Refrigerating and Air-conditioning Engineers*, 2007.