Poster Abstract: Detecting Building Occupancy with Synthetic Environmental Data

Manuel Weber
manuel.weber@hm.edu
Munich University of Applied Sciences
Munich, Germany

Christoph Doblander
christoph.doblander@in.tum.de
Technical University Munich (TUM)
Munich, Germany

Peter Mandl
peter.mandl@hm.edu
Munich University of Applied Sciences
Munich, Germany

ABSTRACT

Information about room-level occupancy is crucial to many building-related tasks, such as building automation or energy performance simulation. Carbon dioxide levels and other indoor environmental factors can be used as a proxy to detect occupancy. In this regard, machine learning solutions have been proposed, with solid performance in detecting presence, as well as counting the number of present occupants, if enough training data is available. The challenge is, to collect sufficient room-specific ground truth data for model training. With this poster, we address the use of knowledge transfer from synthetic data to reduce the amount of required real world data. We outline two approaches for the combination of transfer learning with physical simulations, and motivate the generation of additional synthetic data. Our results show that the required real world training data can be reduced by 50%.

CCS CONCEPTS

• Human-centered computing → Ubiquitous and mobile computing; • Computing methodologies → Machine learning.

KEYWORDS

Synthetic data, CO2, transfer learning, deep neural network

ACM Reference Format:
Manuel Weber, Christoph Doblander, and Peter Mandl. 2020. Poster Abstract: Detecting Building Occupancy with Synthetic Environmental Data. In *The 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '20), November 18–20, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3408308.3431124

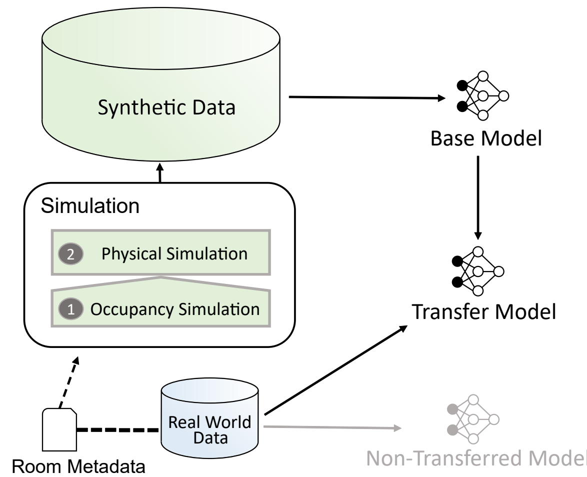

Figure 1: Simulation-aided approach to occupancy detection

1 INTRODUCTION

Human presence is correlated with changes in the indoor climate, in particular the CO_2 level. Although there is a natural lag between human CO_2 generation and sensor perception, it allows to derive occupancy within a reasonable time window, especially for small or medium-sized rooms. Knowing building occupancy on a room-level is essential in many building-related use cases. Application areas include energy performance simulation, building automation or emergency evacuation. Environmental sensing provides several advantages over other sensing technologies. It is more privacy preserving than cameras, and in contrast to light barriers, it does not accumulate errors when occupants enter or leave a room undetected. Machine learning models trained on environmental factors are able to provide solid near-time predictions towards occupancy presence (binary), as well as the count of present occupants [1, 2, 4, 8]. CO_2 was identified as the most important climate factor to detect occupancy [2]. A CO_2-based estimation model can effectively be used in any use case where no down-to-the-minute occupancy data is required, e.g., for annual energy performance simulation.

However, regarding large commercial buildings, the effort of collecting ground truth data in each room impedes applicability in practice. In the same time, simple thresholding solutions are less promising, as measured CO_2 values are influenced by a variety of factors, such as sensor placement, infiltration rates or office plants. Simulated data has been used in other building-related learning tasks, for example in [3], to reduce training times of a reinforcement learning model controlling heating, ventilation, and air conditioning (HVAC) systems. To the best of our knowledge, it has not been studied regarding occupancy detection, which can be used not only in context of HVAC control, but for a variety of applications. Previous studies [1, 7] that apply a transfer of an occupancy model between different rooms show improvements when inferring occupancy from CO_2. Our contribution is a method to lower the amount of required training data for occupancy detection by pretraining with simulated data (see Fig. 1). Further details can be found in [6].

ACM ISBN 978-1-4503-8061-4/20/11.
https://doi.org/10.1145/3408308.3431124

2 SIMULATION-AIDED PREDICTION

Fig. 1 shows our approach on occupancy detection. The idea is to first train a base model on simulated data, and later retrain the model with the given real world data. While the base model can already fit general rules of indoor CO_2 dynamics, the transfer step adjusts the model to room-specific conditions. Taking into account room types and typical occupancy profiles, as well as room metadata such as its dimensions, we conduct two consecutive simulations: (1.) An *occupancy simulation* to generate synthetic ground truth data, including occupancy states for one or multiple occupants, and possibly also occupant behavior (e.g., window opening).
(2.) A *physical simulation*, based on (1.), to generate synthetic CO_2 data, and possibly further environmental data such as temperature or humidity. Room-specific training is still necessary, as conditions differ between rooms, in terms of sensor placement, infiltration rates etc. We distinguish two approaches:

(a) Upfront simulation and pretraining dedicated for the specific room of interest. The simulation process may be automated to complement training procedures.
(b) One-time training of a reusable, general base model. In this case, the base model is dedicated for a group of similar rooms, and simulations are performed for multiple randomly generated room characteristics. Apart from performance improvements with sparse training data, this approach further aims to reduce training times during model application.

In this work, we use approach (a) to conduct a first experiment.

3 RESULTS

We conducted a first experiment in a two-person office room of size 77.5 m³ in a university building in Munich, Germany, to demonstrate the positive effects of transfer learning from synthetic data. We generated 400 days of occupant data for one virtual occupant, and further derived CO_2 values using physical equations described in [5]. Within the simulations several simplifications were made. Infiltration rate, outdoor CO_2 rate and human CO_2 generation rate were held constant, and set to realistic values of 0.0046 m³/s, 360 ppm and 0.24 l/min. Real world data was collected for 7 working days via a Sensirion SCD30 non-dispersive infrared (NDIR) CO_2 sensor and manual recording of occupancy. We applied the deep learning architecture proposed in [4], adjusting it to the case of binary detection from CO_2 only. The model was trained once only on real world data, and once including pretraining. In the latter case, the model was first trained on the simulated data, and model weights were then used for initialisation of the real world training. For evaluation, we applied a cross validation using each day in the dataset, respectively each 2 or 4 consecutive days, for training in one iteration, and the remaining days as test data. Tab. 1 shows the resulting test accuracies and F1 scores for training with 1, 2 and 4 days of training data. It further provides a baseline comparison. Given the scarce amount of data, the model showed clear improvement in the transfer setting: With one day (2 days) of training data, the transfer model exceeded the performance of the non-transferred model trained on 2 days (4 days). Hence, only half of the training data was required. In the same time, model robustness was improved, with standard deviations reduced by up to 50%.

Table 1: Test accuracy and F1 score for the transfer model, non-transferred model and logistic regression (LR) baseline with 1, 2 and 4 days of training data*

		1 Day	2 Days	4 Days
Transfer	Accuracy	0.875 (±0.057)	0.915 (±0.020)	0.933 (±0.015)
	F1 Score	0.764 (±0.171)	0.863 (±0.035)	0.898 (±0.028)
No Transfer	Accuracy	0.715 (±0.086)	0.874 (±0.040)	0.914 (±0.029)
	F1 Score	0.565 (±0.162)	0.791 (±0.065)	0.859 (±0.053)
LR	Accuracy	0.860 (±0.073)	0.879 (±0.058)	0.920 (±0.016)
	F1 Score	0.767 (±0.112)	0.803 (±0.080)	0.866 (±0.038)

*mean values (± standard deviations) are reported for 10 runs in each split

4 CONCLUSION & FUTURE RESEARCH

We propose a simulation-aided approach on occupancy detection in building rooms. Our experiment showed promising results: Required real world data could be reduced by half. The approach has the potential to significantly reduce the time to put a model into production. It also showed remarkable improvement in model robustness. In future work, we intend to find a concrete method concerning transfer as well as simulation of data that is most useful for pretraining. More detail may be added to the simulation, e.g., using historic weather data instead of a constant outdoor CO_2 rate. We are also planning on investigating other model architectures.
The method needs to be validated on a variety of rooms, as it may be more or less beneficial for some room types or occupancy profiles. Considering imperfect air mixing, for large rooms, it can also be necessary to include data from multiple sensors. Moreover, it is particularly interesting to explore the limits of a transfer from simulated data, and to compare results to simulation-free alternatives for data generation, such as methods for time series data augmentation.

ACKNOWLEDGMENTS

Special thanks go to the *OpenPower@TUM* project and the industry partner IBM for providing us with the computing infrastructure.

REFERENCES

[1] I. B. Arief-Ang, F. D. Salim, and M. Hamilton. 2017. DA-HOC: Semi-Supervised Domain Adaptation for Room Occupancy Prediction using CO2 Sensor Data. *4th ACM BuildSys* (2017), 1–10.
[2] L. M. Candanedo and V. Feldheim. 2016. Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models. *Energy and Buildings* 112 (2016), 28–39.
[3] B. Chen, Z. Cai, and M. Bergés. 2019. Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy. *6th ACM BuildSys* (2019), 316–325.
[4] Z. Chen, R. Zhao, Q. Zhu, M. K. Masood, Y. C. Soh, and K. Mao. 2017. Building Occupancy Estimation with Environmental Sensors via CDBLSTM. *IEEE Transactions on Industrial Electronics* 64, 12 (2017), 9549–9559.
[5] P. Parsons. 2014. Determining Infiltration Rates and Predicting Building Occupancy Using CO2 Concentration Curves. *Journal of Energy* (2014), 1–6.
[6] M. Weber, C. Doblander, and P. Mandl. 2020. Towards the Detection of Building Occupancy with Synthetic Environmental Data. arXiv:2010.04209
[7] T. Zhang and O. Ardakanian. 2019. A domain adaptation technique for fine-grained occupancy estimation in commercial buildings. *Proceedings of the 2019 Internet of Things Design and Implementation (IoTDI)* (2019), 148–159.
[8] M. S. Zuraimi, A. Pantazaras, K. A. Chaturvedi, J. J. Yang, K. W. Tham, and S. E. Lee. 2017. Predicting occupancy counts using physical and statistical Co2-based modeling methodologies. *Building and Environment* 123 (2017), 517–528.