

University of Groningen

Office Low-Intrusive Occupancy Detection Based on Power Consumption

Pratama, Azkario Rizky; Blaauw, Frank Johan; Lazovik, Alexander; Aiello, Marco

Published in:
IEEE Access

DOI:
[10.1109/ACCESS.2021.3119997](https://doi.org/10.1109/ACCESS.2021.3119997)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Pratama, A. R., Blaauw, F. J., Lazovik, A., & Aiello, M. (2021). Office Low-Intrusive Occupancy Detection Based on Power Consumption. *IEEE Access*, 9, 141167-141180.
<https://doi.org/10.1109/ACCESS.2021.3119997>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Received September 6, 2021, accepted September 26, 2021, date of publication October 14, 2021, date of current version October 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3119997

Office Low-Intrusive Occupancy Detection Based on Power Consumption

AZKARIO RIZKY PRATAMA¹, FRANK JOHAN BLAAUW^{2,3}, ALEXANDER LAZOVIK³,
AND MARCO AIELLO⁴, (Senior Member, IEEE)

¹Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia

²Research and Innovation, Researchable B.V., 9747 AN Groningen, The Netherlands

³Distributed Systems Group, Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence, University of Groningen, 9712 CP Groningen, The Netherlands

⁴Service Computing Department, Institute for Architecture of Application Systems, University of Stuttgart, 70174 Stuttgart, Germany

Corresponding author: Azkario Rizky Pratama (azkario@ugm.ac.id)

ABSTRACT Precise fine-grained office occupancy detection can be exploited for energy savings in buildings. Based on such information one can optimally regulate lighting and climatization based on the actual presence and absence of users. Conventional approaches are based on movement detection, which are cheap and easy to deploy, but are imprecise and offer coarse information. We propose a power monitoring system as a source of occupancy information. The approach is based on sub-metering at the level of room circuit breakers. The proposed method tackles the problem of indoor office occupancy detection based on statistical approaches, thus contributing to building context awareness which, in turn, is a crucial stepping stone for energy-efficient buildings. The key advantage of the proposed approach is to be low intrusive, especially when compared with image- or tag-based solutions, while still being sufficiently precise in its classification. Such classification is based on nearest neighbors and neural networks machine learning approaches, both in sequential and non-sequential implementations. To test the viability, precision, and saving potential of the proposed approach we deploy in an actual office over several months. We find that the room-level sub-metering can acquire precise, fine-grained occupancy context for up to three people, with averaged kappa measures of 93-95% using either the nearest neighbors or neural networks based approaches.

INDEX TERMS Context-awareness, load disaggregation, occupancy detection, power monitoring, sensor systems and applications, smart metering.

I. INTRODUCTION

Occupancy detection comprises the set of techniques to determine whether a space is empty or occupied, in particular, whether it is occupied by humans and how many of them. Information on occupancy is essential for the automatic regulation of environments in terms of air temperature, air quality, lighting conditions, sound levels, etc. Such automatic regulations are particularly useful in large buildings where they can optimize the energy usage by reducing actuation in unoccupied or less-populated areas [1]. Previous work, including our own, has shown that systems that efficiently plan lighting and heating based on the (expected) occupancy drastically reduce energy consumption (savings have been reported of up to 20%) [2], [3]. Nowadays, many buildings use passive-infrared (PIR) sensors to detect movement as

an indication of presence and control the surrounding lighting and heating systems accordingly. However, PIR-based occupancy detection systems are not ideal, mostly due to their inability to detect motionless occupancy and to differentiate human movement from other movements (e.g., pet movements and wind-blown leaves). Furthermore, PIR-based occupancy detection systems depend on a long set back period (i.e., occupancy timeout), leading to energy wasting.

Recent studies have shown that secondary information (i.e., information sources that are already present for other purposes [4]) can provide a more precise account of occupancy than that provided by PIR sensor data alone. For example, previous studies perform occupancy detection by using room temperature and CO₂ concentration measurements [5] or by looking at power consumption [6], [7]. The latter is of interest mainly for the following three reasons: (i) occupancy state changes can be measured immediately, (ii) electricity is less affected by invisible environment noise, such

The associate editor coordinating the review of this manuscript and approving it for publication was Chun-Wei Tsai¹.

as interference of radio frequency (RF) signal transmission and unseen factors that impinge on CO₂ and temperature readings, (iii) power meters are affordable and are becoming ubiquitous in buildings.

Although various studies have investigated the use of power meters for occupancy detection, they offer different level of granularity. In the context of the present work, granularity can be understood as the ratio between the size of the object of interest and the minimum space it can be detected in. For example, if one is only able to detect the presence of a person in a household [8], [9], one would say that the detection method is coarse, while being able to detect a person working at a specific desk is fine grained. If one is further able to detect even the number of people present and assign an identity to them, one considers the system to be even finer grained. The finer granular methods increase the amount of energy that can be saved by personalizing the heating, ventilation, and air conditioning (HVAC) system based occupants' profiles. To attach more details, however, more power meters are required to observe specific persons or particular devices [7], [10]–[12]. Deploying a power meter for each device or for every person is most often unfeasible due to deployment- and maintenance-costs, in addition to being privacy intruding. Therefore, we explore using a limited number of power meters and consider the following research question: “How reliable and fine grained can occupancy detection be by deploying power meters per-room or per-area?”

To answer this question, we explore the retrieval of fine-grained occupancy information in shared offices by using aggregate power measurements and by applying machine learning techniques. We hypothesize that fine-grained occupancy information can be extracted from aggregate room- or floor-level power measurements. That is, we only consider the availability of one power meter per circuit breaker covering several workspaces. By focusing on the aggregate power usage we take the first steps towards non-intrusive occupancy detection for reducing energy consumption. We apply machine learning algorithms to analyze the aggregated power consumption and to investigate the occupancy by using both a sequential (i.e., time series) and a non-sequential (i.e., cross-sectional) approach, and compare their performance.

Our contributions are:

- 1) A fine-grain occupancy detection in a shared office based on aggregate power consumption
- 2) An evaluation of classification techniques for the specific problem at hand that is particularly suited to improve the inference results.

The paper is organized as follows. Section II overviews related work covering non-intrusive load monitoring (NILM) studies and data mining on power consumption. In Section III, we describe our power consumption based occupancy detection approach. We describe the acquired data collected during experiments and evaluate the approach using different classification techniques in Section IV. In Section V, we discuss

the proposed occupancy detection system and review proposals with other sensing modalities as a comparison. Finally, Section VI summarizes and discusses the findings presented in the paper.

II. RELATED WORK

Power meters are ubiquitous in all households and buildings connected to the power grid. Besides their main purpose of measuring power consumption for billing, power meters have also been used for other applications, such as the recognition of power-consuming devices [13], [14] and the inference of the occupancy state of buildings [6], [8], [9]. The former is known as *Non-intrusive Load Monitoring* [15], while the latter is related to *Power consumption data mining* [11]. Next we overview related work in these two areas.

A. NON-INTRUSIVE LOAD MONITORING

Non-intrusive Load Monitoring (NILM) is a technique that allows for the identification of the state of an individual device without explicitly using a dedicated power meter to monitor its behavior. The term was first coined by Hart, *et al.* [15]. NILM typically aims to decompose total energy usage per appliance and to detect power hungry or faulty appliances.

A large part of research in NILM focuses on the residential sector and uses publicly available datasets. Basu, *et al.* used the REMODECE dataset,¹ which contains data of 100 households with up to a year of measurements per household [14]. The authors approached the disaggregation problem by using a time series based classification approach. To do so, they first segmented the measurement into various sub-sequences, after which they assigned a class label for each sub-sequence, based on the closest distance to the class label in the training data. The authors reported that using the Euclidean distance and dynamic time warping (DTW) as distance metrics outperformed the temporal correlation (TC) metric. The nearest neighbor method (using these three metrics) performed better than the Hidden Markov Model (HMM). In all houses, Water heaters were recognized quite well, achieving an F1-score of up to .94 in 10-fold cross-validation. Similarly, Cominola, *et al.* used the Almanac of Minutely Power dataset [16] (an open residential dataset) in their experiment [13]. They applied a hybrid approach by combining Factorial HMM (FHMM) and DTW for the correction of inaccurate classification outputs. For high-power appliances, such as heat pumps, refrigerators, network security equipment, and HVAC loads, they reported an F1-score over .95.

Ruzzelli, *et al.* performed their classification using a non-linear model (neural network) approach [17]. The authors proposed *RECAP*, a framework for online load disaggregation. RECAP focuses on discriminating two appliances that have similar or the same energy consumption. They reported an accuracy of .84 for recognizing room heating

¹ See <https://remodece.isr.uc.pt> for more details.

devices, water heaters, microwave ovens, and refrigerators, each consuming over 2,000 Watts.

Akshay *et al.* used occupant location as an additional input in power consumption disaggregation [18]. Their framework, *LocED*, derives user occupancy from Bluetooth low energy (BLE) beacons and WiFi access points, by determining a person's location based on nearby devices. The authors used the location as a basis and determined a set of appliances that might be being used by occupants by assuming that the position estimation is reliable. A combinatorial optimization (CO) algorithm was then used to segregate power consumption by finding the combination of appliances that produce a total consumption closest to the current measurement. While we also aim to exploit the aggregated power consumption, we utilize electricity consumption measurements to infer occupancy in contrary to the BLE and WiFi approach used in *LocED*.

Kelly *et al.* applied deep neural networks to separate the consumption of individual devices [19]. They used the open dataset UK-DALE [20] in their experiment. The dataset consists of measurements in a home with five devices with high power demand (300 – 3,100 Watts). The authors compared three different neural network architectures, including recurrent neural networks (RNN) with long short-term memory (LSTM) cells. They trained several network models, each representing one target appliance on one complete cycle of activation. As the target appliances had different activation durations, the window width varied based on appliance type (the window ranged from 128 samples/13 minutes to 1,536 samples/2.5 hours). The results showed that the three RNN models outperformed both combinatorial optimization and FHMM. Among the three networks, LSTM achieved the best results for two-state appliances.

When shifting focus from a residential environment towards a public environment, appliance recognition using one sensor measurements becomes challenging. There might be numerous identical devices running concurrently, preventing any useful context extraction from data. As a solution, Zoha *et al.* argued that more power meters need to be deployed, for example in circuit breakers, to improve the quality of gathered information [21]. They performed appliance recognition using an FHMM to recognize several appliances commonly found in a working environment, such as workstations, monitor screens, laptops, lamps, and table fans. They reported an F1-score of up to .90 in on/off devices and an F1-score of up to .80 in multi-state devices.

Most often, NILM studies involve high power devices and very rarely take low power devices into account. Furthermore, the aforementioned works are only concerned with specific appliance recognition, and did not cover occupancy inference. Our previous work addressed low power devices recognition [22] and monitor screen activation detection [23]. In this work, we move one step forward with occupancy inference.

B. POWER CONSUMPTION DATA MINING

Power consumption analysis in learning occupancy approaches mainly address occupancy detection in homes or workspaces to enhance energy awareness and optimize energy usage.

1) HOME OCCUPANCY

Several studies revealed that there is a (strong) correlation between power consumption and home occupancy. Kleiminger *et al.* observed five houses over an eight-month period, during which they used a central power meter to measure the consumption of specific appliances [8]. Moreover, they also deployed several power meters to some devices and a PIR sensor near the doorway for further analysis. The authors report an accuracy higher than .80 in most scenarios (applying *k*-NN and HMM techniques). Similarly, Dong Chen *et al.* investigated the potential of using smart meter data for home occupancy detection [9]. They observed two houses, each with separated circuits. The main circuit was designed for supplying electricity to background loads only, such as a refrigerator and air conditioners. Other devices whose activation indicates physical interaction with occupants (for instance, a microwave and wall switches) were connected to branch circuits. To generate occupancy traces, they detected events of any power changes (i.e., with more than a 30 Watts difference) followed by clustering the nearby events.

2) ROOM AND WORKSPACE OCCUPANCY

Other research has addressed occupancy inference in offices with extensive measurement units (i.e., one power meter per device) [10], [11]. Shetty *et al.* observed the presence of four workers by monitoring the consumption of monitor screens and PCs, any by using a PIR sensor in each workplace [10]. To assign presence and absence states, they performed a *k*-means clustering analysis on the measurement data. They reported highly accurate occupancy inference, reaching an accuracy of .98. Zhao *et al.* observed more detailed user behavior by extensively measuring plug loads and categorizing the loads into three classes: personal computers, lighting, and other appliances [11]. The authors used per user measurement data to train several models of machine learning, including decision trees, support vector machines, and naive Bayes classifiers. Their aim was to detect occupancy and computer activation states in office environments. Furthermore, room occupancy levels were also approximated using a regression analysis algorithm. It was reported that decision trees performed best for recognizing ten participants, reaching an average accuracy of .90 and a Kappa value of .69. For occupancy level prediction, the comparison between the prediction and the ground truth showed a strong correlation, reaching .95. While the reported results are good, the initial cost to invest scales with the number of appliances to be monitored.

Petrovic *et al.* addressed occupancy inference by observing WiFi router power consumption [6]. They developed a power meter based on an Arduino microcontroller to measure the power consumption of a router at a relatively high sampling rate, i.e., 0.5 Hz. In addition, they monitored twenty commercial plug-level power meters to monitor office appliance activities and infer occupants. The power measurements are then benchmarked against the occupancy inference based on WiFi power-consumption. To infer occupancy, the authors extracted several features from a moving window of sensor readings and applied a random forest classification algorithm. They found a correlation between the number of occupants and the increase in power consumption of the router. Their occupancy detection algorithm had an accuracy of approximately .93. The accuracy improved by .03 by taking plug-level power measurements into account.

While high power consumption may correlate with occupancy, one still needs to extract information from power readings. Most studies, however, have not treated sub-metering in an office room. In the following section, we elaborate on how we deal with the aggregate power readings measured at a circuit breaker of an office room.

III. METHODOLOGY

We propose a generic approach to infer occupancy using sub-metering power consumption in an office room. To support our approach, we initially observed individual consumption to understand how users consumed energy during their occupancy periods. With this information, we developed multiple machine learning models and performed occupancy context classification on the aggregated power of the room.

A. DATA COLLECTION

Power measurements were collected using *Smappee*² sensors: a three-phase clamp-based power meter. The installation was performed by placing clamping clips on the electricity sources. We used two Smappee power meters yielding a total of six available clamps. One of the clamps was dedicated to measuring the total power consumption of all participants in a room, while the other clamps were attached to each power plug (i.e., one clamp per user). With this setup, we collected individual and total consumption for about two months. Based on the measured data, we built user energy profiles that provide information about the consumption pattern of the individual that contribute to the aggregated consumption. For example, these profiles represent how much power is consumed and how the temporal patterns are drawn [24]. Using multiple Smappee meters also speeds up the process of collecting labeled training data by concurrently collecting different combinations of individual consumption. In an actual deployment, we only need one clamp to measure aggregate power consumption in an electric phase, for example, clamped at an input circuit-breaker line in a room- or floor-level. To supervise machine learning models, we can

utilize the other two clamps to alternately collect labeled training data.

In order to store the consumption for offline analysis, we developed a custom gateway application that forwards the sensor data to our data warehouse. The gateway is sensor specific application that transforms the specific Smappee data format to a generic time series data format.

Each measurement in the time series consists of five variables: (i) active power, (ii) reactive power, (iii) apparent power, (iv) power factor, and (v) electric current. Active power is the total amount of power that flows through the Smappee clamp, and is consumed by the devices and electrical resistance (measured in Watts). Reactive power is the dissipated power as a result from inductive or capacitive components in appliances (measured in Volt-Amps-Reactive, VAR). Apparent power is the product of the root-mean-square voltage and the root-mean-square current (measured in Volt-Amps, VA). Power factor or cosphi (in percents) represents the ratio of the active power flowing to the appliance divided by the reactive power. Electric current is the amount of electron flows through the clamp (measured in ampere). All these variables were collected in a five-second interval. We downsampled this signal into a 1-minute interval to reduce computation load.

The resulting data are a time series, which means that each observation consists of measurements collected in chronological order to form a sequence. The resulting observations might be incomplete, for example, when failures occur during the measurement. When data is missing, we assume that there is no value change and use last observation carried forward imputation (up to at most 10 minutes). The actual, ground-truth participant occupancy was manually collected in a spreadsheet document. As the ground truth might be incomplete during a longer period of data collection (e.g., when the observer was not present), we generated class labels to refine ground truth by filling the missing annotated labels. This was done by applying a threshold to the per individual power consumption. The reason for this is that occupant presence is indicated by the interaction with appliances and can be recognized by a change in power of a certain amount [9]. We chose 20 Watts as a threshold based on empirical observation on monitor screen activation. To validate how the threshold method fits with the real occupancy, we deployed a webcam and captured occupancy images on a two minute interval during a week (see Section IV-B).

B. OCCUPANCY DETECTION

Given a set of individuals $J = \{j_1, j_2, \dots, j_n\}$ and a set of electricity features $M = \{m_1, m_2, \dots, m_h\}$, the aggregate power reading at time t is

$$\vec{X}_t = [f_1(\vec{x}_{t,m_1}), f_2(\vec{x}_{t,m_2}), \dots, f_h(\vec{x}_{t,m_h})], \quad (1)$$

where $\vec{x}_{t,m_k} = [x_{t,m_k}^{j_1}, \dots, x_{t,m_k}^{j_n}]$ is the vector of feature readings m_k of individual $j_i \in J$, and $f_k(\vec{x}_{t,m_k})$ is the aggregate summary function of feature m_k over all individuals $j_i \in J$.

²Website: <https://smappee.com>

TABLE 1. Class labels and meanings.

Labels	Y_t	Context
0	[0,0,0]	Nobody is around
1	[1,0,0]	$y_t^{j_1}$ is present
2	[0,1,0]	$y_t^{j_2}$ is present
3	[1,1,0]	$y_t^{j_1}$ and $y_t^{j_2}$ are present
4	[0,0,1]	$y_t^{j_3}$ is present
5	[1,0,1]	$y_t^{j_1}$ and $y_t^{j_3}$ are present
6	[0,1,1]	$y_t^{j_2}$ and $y_t^{j_3}$ are present
7	[1,1,1]	all individuals are present

An example of feature is active power a which has an aggregation function of active power $f_a(x_t, a) = \sum_{i=1}^n x_t^{j_i}$.

The occupancy detection function f_{occ} assigns Y_t , the summary of presence states of all individuals, to the power reading \bar{X}_t , that is, $f_{occ}(\bar{X}_t) = Y_t$. Y_t is a class label that represents the presence state of all individuals and can be transformed into the binary occupancy state of n individuals $Y_t = [y_t^{j_1}, y_t^{j_2}, \dots, y_t^{j_n}]$, where $y_t^{j_i} \in \{0, 1\}$. We aim to develop estimators that predict the occupancy state $y_t^{j_i}$ for a specific individual j_i based on aggregated measure readings \bar{X}_t . Hence, the number of classes will be 2^n , where each class represents different occupancy states of individuals (i.e., *absence* or *presence*). Table 1 illustrates the class labels and their representation. Each class label identifies the exact occupants present in the shared office.

1) GENERAL APPROACHES

We divide our data set into a training, validation, and test set. We investigate two different directions for creating the training and validation sets in order to indicate how occupancy classification results differ by the way the data is divided. Firstly, a subset of data is selected based on a randomized shuffling by preserving the proportions of the class prior probabilities (i.e., to represent the overall data distribution). This division scheme reduces variance and ensures the models' generality. Secondly, we divide the data based on the historical occurrence and assign the last series of days in the dataset as a test set. The same data division scheme was also done in related previous work, e.g., [25]–[28]. The purpose of this data division is to see the performance over the last few days of the collected dataset that represent a condition when the system is deployed in a real-world and has no access to retrain the classification models with the unseen training data. We refer to this detection scheme as *daily occupancy detection* in the following sections.

Our model fitting procedure takes the form of a three step approach. In the first step, we use the training and validation sets (85% of the total data set) for training classification models and to find the optimal model parameters. In order to make as much use of these data as possible, we apply a 5-fold cross-validation procedure. Cross-validation is a well-known procedure in the field of machine learning, in which all data in the training set is used for training an instance of the model four times, and used as validation (i.e., not used for training the model) once [29]. With cross validation, as much data

as possible is used to train the model, whilst still leaving room for validating the model on unseen data. After fitting the initial model, the next step consists of using the parameters of the model and retraining the model using the combined training and validation set as the full training set. In the final step, the validation step, the final model is used to classify the remaining 15% of the data to evaluate its unbiased performance.

We applied various preprocessing steps to improve the models' performance. Firstly, We applied a normalization procedure to scale values between 0 and 1. This step is necessary as the ranges of the measurements fluctuate heavily. For example, the active power of an individual may exceed 120 Watts while the current readings are below 1 Ampere. Furthermore, normalization helps the neural network and nearest neighbor classifiers to better determine the decision boundary.

Secondly, we performed relabeling in the sequential data. In estimators that deal with sequential data, often a sampling window does not fully represent a single class (e.g., state changes might happen in the middle of a window period). When this is the case, it is difficult to provide scrupulous outlier-free training data to classifiers. State changing in the middle of a sequence with a particular label could negatively influence the learning process of the sequence, and thus be considered as an outlier. We perform preprocessing on the training data by taking only full-length and partially complete sequences that represent one class. Full-length sequences refer to the L -consecutive instances with the same label within an L -sized window. The partially complete sequence is a series of instances with a homogeneous label, but where the length is slightly less than L . These sequences are illustrated in Figure 1. For the latter, we replicate the last value and impute the replication to form full length L -consecutive instances.

Thirdly, we performed feature engineering. We investigated a different combination of raw time series variables to discover the potential patterns formed during occupancy. This includes (i) Watts, VAR, VA, current, and cosphi, as in [7], (ii) Watts only, as the most basic measurement component in a power meter, (iii) VAR and Watts, as proposed by Hart *et al.* [15], and (iv) Watts, current, and cosphi. We also add features that indicate the range of the time of day when a measured value occurred. We mark the instances as a member of the corresponding time of day. The markers, represented using one-hot-encoding, are considered as additional features to the estimators.

After the preprocessing steps, we performed hyper-parameter optimization to find a set of hyper-parameters (or tuning parameters) that work best on the training and validation sets. Examples of hyper-parameters are the size of sequence length L , the number of k -neighbors for nearest neighbors, and the number of neurons for neural network based techniques.

We finally evaluate the performance on the test set using models that are retrained to the combination

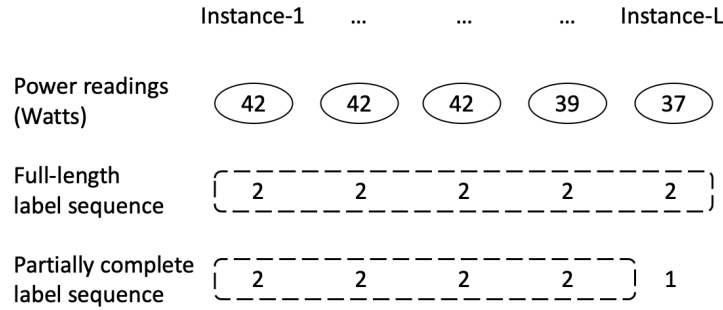


FIGURE 1. Full-length and partially complete sequences representing class label 2.

of the training set and validation set using the best hyper-parameters.

2) METHODS

We apply several state of the art machine learning techniques, in particular, pattern classification algorithms [30]. We implemented and compared several algorithms to perform occupancy classification. For this, we made a distinction between (i) sequential, (ii) non-sequential, and (iii) generative algorithms. The sequential algorithms take the time ordering of the data into account, and can also consider lagged versions of features. The non-sequential algorithms assume that the specified features are independent of each other, and train a regular machine learning algorithm only based on contemporaneous data. Finally, we considered generative classification which model the underlying distributions of the classes.

For both sequential and non-sequential classification we used and compared (the adapted versions of) the k -Nearest Neighbor (k -NN) algorithm and the neural networks algorithm. The k -NN based algorithm predicts class labels based on the majority vote to unlabeled query data [31]. The neural networks approach fits a nonlinear estimator for regression or classification [32]. The model initially derives hidden features from the inputs followed by modeling classification as a function of the combination of the hidden features.

a: SEQUENTIAL CLASSIFICATION

Sequential classifiers take historical data into account. Sequential classifiers only have an advantage when a variable x_t depends on past observations of x , $x_t \not\perp (x' \in \{x_{t-1}, \dots, x_0\})$. In this project, an assumption is made that once an individual is present and consuming electricity, the devices he or she uses will stay activated for a longer period of time, causing some sort of serial correlation.

We adapted the k -NN algorithm to do sequential classification (i.e., k -NN^{seq}) by appending lagged measurements to form the L -consecutive instances of M -dimensional feature vector of measurements. The $L \times M$ part of measurement is assigned to a single class. The sequencing means that before a classification is being done, the full sequence of feature vectors needs to be prepared. The k -NN classifier then

compares the distance between the query sequence and train data sequences.

In the sequential classification approach, we extend the traditional, feedforward neural network architecture by considering recurring events, that is, we apply the Recurrent Neural Network (RNN) approach. Apart from contemporaneous input values, RNNs also consider the previous input values in order to predict outcomes. We use a sliding window approach to create sub-sequences of our time series data. In our previous research, we found that the best performance is achieved when these windows are non-overlapping, hence the decision to apply the same methodology here [33].

b: NON-SEQUENTIAL CLASSIFICATION

In the non-sequential classification case, we applied pattern classification techniques. Given a power reading \vec{X} as a vector input, the models should learn how to map the readings to a single class label $y^i \forall j_i \in J$. Note the lack of the subscript t , indicating that these predictions are considered independent of the time or preceding measurements.

As in the sequential classification case, we apply both k -NN and neural networks, as their discriminating performance often performs well in this setting [8], [14], [17]. For both k -NN and neural networks we used the same number of input vectors and output classes.

c: GENERATIVE CLASSIFICATION

Finally, we applied a generative classification method. This approach learns models that generate data and use the models to classify instances. The occupancy based on power consumption can be regarded as a Hidden Markov Model (HMM) problem as the occupancy state cannot be observed directly. The occupancy is only indicated by observable power energy that he or she consumed. This way, we construct an HMM chain for each individual.

As we use a power meter that measures aggregated power consumption, the model is generalized to a Factorial HMM [34]. Figure 2 shows the illustration of the FHMM chain, where the aggregated power reading X_t is affected by the unobservable presence state of individuals $y_t^i \forall j_i \in J$.

The exact computation of FHMM can be implemented as the cross product of the state variable of each individual. This computation forms an equivalent HMM chain in which each

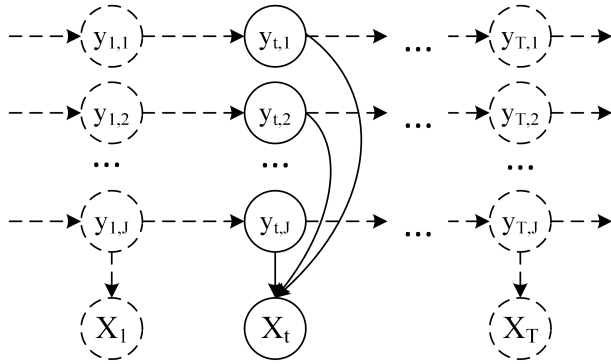


FIGURE 2. A factorial hidden Markov Model chain.

state represents one combination of the employees' occupancy states. Although this approach grows exponentially with the number of occupants, it is still tractable for a few participants, as in this work. The optimal sequence of hidden states can be estimated from the FHMM using the Viterbi algorithm [35].

C. METRICS

In order to evaluate the classification, we define the following metrics:

1) OVERALL ACCURACY

Accuracy shows the classification performance by calculating the number of correctly predicted labels divided by the total instances that have been classified, as given in Eq. 2.

$$\text{actualAccuracy} = \frac{\text{correctly predicted class}}{\text{total instances}} \quad (2)$$

2) KAPPA MEASURE

We consider Cohen's Kappa measure to avoid bias in indicating the accuracy. This measure is defined as the agreement between the actual predictor and the random predictor, as shown in Eq. 3 [36].

$$\text{kappa} = \frac{\text{actualAccuracy} - \text{randomAccuracy}}{1 - \text{randomAccuracy}} \quad (3)$$

The actual accuracy is the success rate of actual predictor, while the random accuracy is the success rate of random predictor, that is, a hypothetical expected probability of agreement under an appropriate set of baseline constraints [36]. Mathematically,

$$\text{randomAccuracy} = \frac{\sum_{c \in C} \text{actualClass}_c * \text{predictedClass}_c}{\text{totalInstances}^2} \quad (4)$$

3) RECALL AND PRECISION

For two-class classification, such as in the classification of j_i 's binary presence state, we provide recall and precision calculation. Recall shows how good the classifier is in detecting the individual presence over the overall presence condition, while precision shows how good the correctly predicted presence is

over the overall prediction of being present.

$$\text{Precision}_{j_i} = \frac{TP_{j_i}}{TP_{j_i} + FP_{j_i}} \quad (5)$$

$$\text{Recall}_{j_i} = \frac{TP_{j_i}}{TP_{j_i} + FN_{j_i}} \quad (6)$$

TP_{j_i}/TN_{j_i} counts the number of instances for which j_i 's presence/absence are correctly predicted. Correspondingly, FP_{j_i}/FN_{j_i} counts the number of instances for which j_i 's presence/absence are misclassified.

IV. EXPERIMENTAL EVALUATION

Participants were four male PhD students of the University of Groningen, Netherlands. The students had individual desks in a shared office room. During the study, we collected two data sets of aggregated power consumption. The first dataset we collected contained the aggregated power consumption data which was used for training and testing the estimators. We split this data set into a training and test set with randomization and stratification (see Section IV-C) and one without randomization (see Section IV-D). The data was collected during a two-month period between the 31st May 2018 and the 11th September 2018. The second data set comprised aggregated power consumption data, but this time it was combined with camera recordings, to verify the quality of the ground truth data. The latter data set was collected from the 4th February until the 8th February 2019. Note that one of the participants was only included in the first part of the experiment, and replaced with another participant in the second part of the experiment. As such, the devices and user behavior of the new participant differ from his predecessor. All participants provided written informed consent for their data to be used in scientific research.

A. EXPERIMENTAL SETUP AND DATA COLLECTION

The observations comprise 50 weekdays of data. The days are uniformly distributed, as shown in Figure 3, to ensure natural energy consumption of each participant is captured. We only consider work hours from 7.00 AM to 9.00 PM to avoid performance bias due to higher score of correctness in detection of person being vacant outside of typical work hours.

The observed participants did not follow a strict regular work schedule and could start with working at any time. One of the participants (P1) was also present during the weekend. Figure 4 shows the proportion of the subjects' presence during the observation period. The colors of the outer ring represent participants, whereas the color of the inner rings depict the days when they were present. The size of the sections of the rings depict the respective proportion of total presence. It can be seen that participant P3 was the most frequently present with 35 days, followed by P1 and P2 with 29 and 19 days, respectively. We found that on average on Tuesday most people were present. P1 and P2 were less frequently present on Mondays and Wednesdays; P3 was less frequently present on Thursdays.

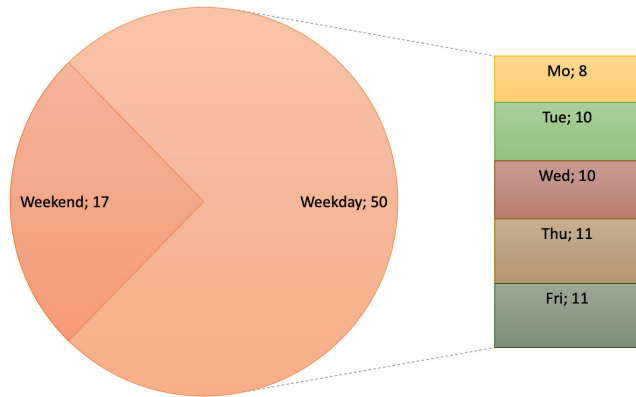


FIGURE 3. The weekdays and weekends distribution in the dataset.

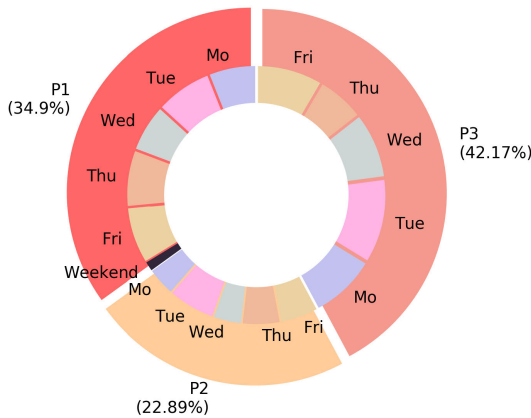


FIGURE 4. The portion of each individual presence.

The probabilities of a participant being present is shown in Figure 5. The participants generally came in at 10.00 AM. The presence of participant P1 showed stronger variation than the presence of the others, while P3 consistently came in at around 10.00 AM. As for departing time, P1 and P3 were leaving at around 6.00 PM and 4.00 PM, more than half of the times, respectively. Of the three persons being observed, P3 was the most consistent participant in terms of work hours, that is, usually present from 10.00 AM to 4.00 PM.

Figure 6 shows the probability distribution of the total power consumption. It can be seen from the figure that P1 and P2 consumed in average 52.50W ($\sigma_{\bar{x}} = 9.90$) and 130.18W ($\sigma_{\bar{x}} = 18.18$) during the work-related presence, respectively. Interestingly, P3 has two notable peaks, that is, at around 29 and 39 Watts during his work sessions which might be due to two forms of frequently used group of appliances (for example either laptop or desktop PC interchangeably).

B. VALIDATION OF OCCUPANCY DETERMINED BY THRESHOLDING

Table 2 shows a comparison of occupancy prediction based on individual energy consumption with the camera-based ground truth. For this, we used a threshold on the energy consumption to determine whether participants were present (i.e., the power consumption had to be more than 20W).

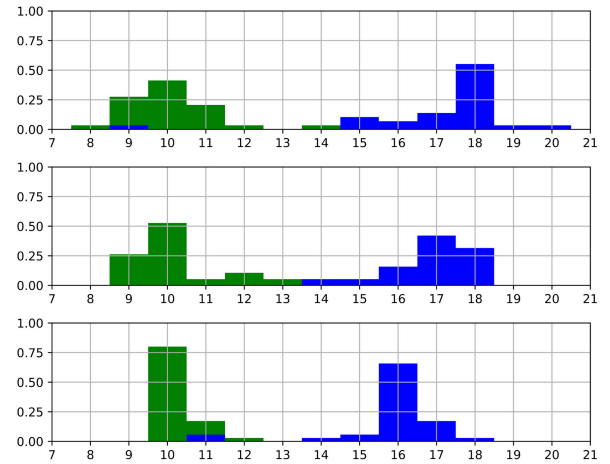


FIGURE 5. The time-dependent presence probability of coming to (green bars) and leaving from (blue bars) the office. From top to bottom: P1, P2, and P3.

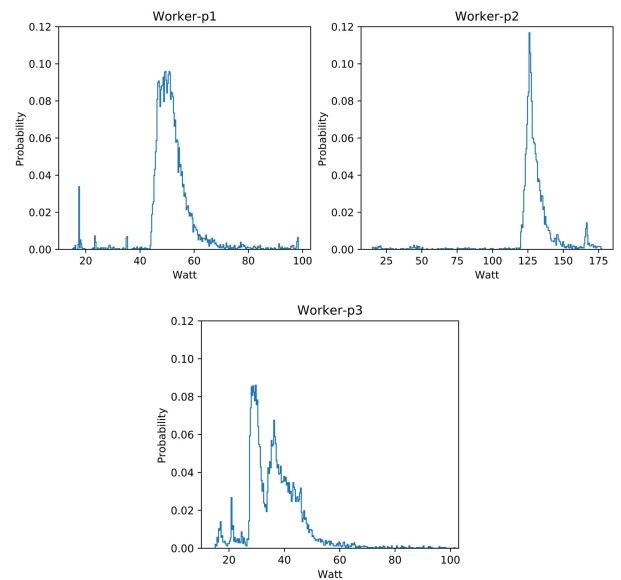


FIGURE 6. The Power distribution of the three observed participants: P1, P2, and P3.

This comparison is necessary to evaluate the reliability of the thresholding method. The comprehensive labels are then useful in training classification models.

The table shows that two of three participants (i.e., P1 and P3) had the kappa measure of agreement of 0.97 between the occupancy based on the power threshold and the occupancy based on the camera observation. The precision and recall ranged between .967 and .983, respectively, for both individuals. Unfortunately, at the time of camera deployment, P2 had left and was replaced with P4. P4 had the kappa measure of agreement of .64, while the recall and precision were 0.628 and .806, respectively.

C. OCCUPANCY DETECTION

We used the aggregated energy readings \bar{X}_t to predict the occupancy state $y_t^{j_i}$ for each specific individual j_i . We applied

TABLE 2. Comparison occupancy prediction based on individual energy consumption to the camera-based ground truth. P2 has no camera recordings available as he has moved during the camera deployment period.

Subject	Acc	Kappa	Recall	Precision
p1	.988	.967	.967	.983
p3	.988	.969	.972	.982
p4	.888	.638	.628	.806

TABLE 3. Averaged precision of 5-cross validation classification based on all features (watts, VAR, VA, current, cosphi) on 85% shuffled train validation set, investigated on various sequence length.

Sequence	1-min	2-min	5-min	10-min	20-min
RNN	0.931/0.873	0.954/0.915	0.949/0.906	0.943/0.895	0.937/0.885
k -NN ^{seq}	0.969/0.943	0.967/0.940	0.959/0.925	0.952/0.913	0.941/0.893

machine learning approaches in which we investigated the classification of sequences with different length and feature combinations. We followed the approach mentioned in Section III-B1 and applied randomized shuffling for training validation set division. By shuffling this way, the process keeps the proportion of class representation as in the whole dataset.

Table 3 shows how sequence length affects model performance for classification using both RNN and k -NN^{seq} on the shuffled training validation set. Generally, the k -NN^{seq} model performs better than the RNN model when using the shuffled training and validation sets. Using RNN, the average classification accuracy and kappa measure on 1-minute sequence lengths are .931 and .873. The precision increases when we increase the sequence length, reaching the optimal performance in 2-minute sequence lengths with .954 accuracy and a kappa of .915. The precision is then going down as we increase the sequence length. Interestingly, this trend is different in the sequence classification using the k -NN based algorithm. The k -NN^{seq} model performs optimally when using 1-minute sequences, reaching .969 accuracy and a kappa of .943. The performance degraded when we considered longer sequences, reaching an eventual .941 accuracy and a kappa of .893 in 20-minute sequences.

To get a better understanding of sequence-based analysis on different feature combinations, we set a 2-minute sequence length for both the k -NN^{seq} and RNN algorithms. Table 4 presents the best average performance of 5-fold cross-validation on training and validation sets. Active power (Watts), which is the most common electrical variable of a smart meter, can reveal occupancy agreements ranging from a kappa of .881 to .902. The performance of the models gradually improved, reaching the accuracy of .954 to .967 and the kappa of .915 to .94 when all the five measured variables were included (showed as set-4). The maximum performance of the models was reached when temporal features were added as predictors (showed as set-5), reaching a kappa of .926 and .951 for RNN and k -NN^{seq}, respectively. We saved the parameters of the best models indicated by the highest precision metrics, then used them in retraining the same model using the whole training validation set (85% of total data) and test the 15% new data.

TABLE 4. Averaged precision of 5-cross validation classification based on 2-minute sequence length on 85% train validation set, investigated on various feature sets.

Set-	Features	RNN		k -NN ^{seq}	
		Acc	Kappa	Acc	Kappa
1	Watts	0.936	0.881	0.947	0.902
2	VAR and Watts	0.941	0.890	0.963	0.932
3	Watts,current, and cosphi	0.942	0.893	0.966	0.938
4	Watts,VAR,VA,current,cosphi	0.954	0.915	0.967	0.940
5	Set-4 + time of the day	0.960	0.926	0.973	0.951

TABLE 5. Network configuration for NN and RNN.

Parameters	NN	RNN
Memory units	-	100
Neurons	30	-
Activation function	Rectified Linear Unit	LSTM cells
Weight optimizer	Stochastic Gradient descent	Adam optimizer
Cost function	cross entropy	cross entropy

Using the saved parameters, we classified the test set and presented the results in Table 6. As for k -NN algorithms, the optimal parameter for both sequence and non-sequence were $k = 11$. In the basic feed forward neural network, we found that 30 neurons was the optimum for the model's performance. We applied early stopping and a maximum of 800 iterations and used stochastic gradient descent for the weight optimizer and rectified linear unit (ReLU) activation optimizer. We repeated the experiments for five times and took the average of the accuracy and kappa measure. In the RNN, we apply LSTM cells as activation function with Adam Optimizer [37]. We apply 100 hidden units in a single cell in a LSTM layer and using 744 epochs. We have also investigated multiple LSTM layers, however, it did not deliver a significant improvement. The network configuration is summarized in Table 5.

The performance of k -NN achieved an accuracy of .966, a kappa of .937, and a F-measure of .910 when evaluated in the non-sequential form. There were no significant changes when evaluated in the sequence form. The averaged performance of the basic neural network, as measured by the accuracy, kappa-statistic, and F-measure were .941, .891, and .835, respectively. We can expect about 5% of kappa measure improvement when we classify the same problem with 2-minute length sequences using RNN. The classification using FHMM had the worst performance, reaching an accuracy of .774, a kappa of .622, and F-measure of .428.

D. DAILY OCCUPANCY DETECTION

In the second experiment, we used the same dataset and splitting portion as discussed in Section IV-C, but we did not shuffle the dataset. That is, for this experiment we divided training and validation sets based on historical occurrence. We took the first 85% portion of the dataset and used that for training validation, and left the remain portion (where the test set starts from and includes the 3rd September 2019).

With this configuration, we focused only on the sequence classification (e.g., using k -NN^{seq} and RNN) because of their higher performance than the other non-sequence based classification (see Table 6). The best parameters of the RNN in the

TABLE 6. Classifier accuracy and Kappa measure (with hold out test set, train-test-shuffle = True).

no	Techniques	how to model	k-cross-validation	acc	kappa	F-measure
1	k -NN	1-nearest neighbor classifier	10-fold-cv	0.966	0.937	0.910
2	k -NN ^{seq}	1-nearest neighbor classifier	GroupKfold	0.970	0.946	0.916
3	NeuralNet	1-network architecture	10-fold-cv	0.941	0.891	0.835
4	RNN	1-network architecture	GroupKfold	0.964	0.934	0.905
5	FHMM	n -Markov Chains	no cv	0.774	0.622	0.428

TABLE 7. Occupancy prediction as measured by the Kappa statistic, based on 2-minute length sequences.

Date	Max persons	Occupancy with ID		Occupant counting	
		k -NN ^{seq}	RNN	k -NN ^{seq}	RNN
2018-9-3	3	0.865	0.841	0.911	0.858
2018-9-4	3	0.877	0.801	0.905	0.816
2018-9-5	1	0.850	0.876	0.855	0.876
2018-9-6	2	0.559	0.628	0.737	0.796
2018-9-7	3	0.787	0.897	0.783	0.911
2018-9-10	3	0.717	0.836	0.859	0.888
2018-9-11	3	0.854	0.952	0.855	0.960
7-day average		0.787	0.833	0.844	0.872

non-shuffled set were different from the shuffled one. That is, we achieved the best results using 65 hidden neurons with a single cell LSTM layer within 53 epochs. As for the k -NN^{seq}, we found that $k = 11$ nearest neighbors delivered the best results.

As there was no shuffling process, the dataset is naturally sequenced in historical order. We can thus provide occupancy detection on a daily basis as presented in Table 7. We present the agreement of occupancy detection and actual occupancy as measured by the kappa statistic. We also show the fine-grained occupancy inference (i.e., presence state with ID) and occupant counting (i.e., the estimation of the number of occupants).

As shown in Table 7, classification based on sequence using RNN generally achieved higher kappa measures than the non-sequence classification using k -NN^{seq} on non-shuffled data set (i.e., data from the 3rd of September 2018 until the 11th of September 2018). The classification accuracy using the RNN can supersede the accuracy of the k -NN^{seq} with up-to 12% for occupancy detection with ID, reaching a kappa of .836 on the 10th of September 2018.

A comparison of the two algorithms on the 10th of September 2018 is illustrated in Figure 7 and Figure 8. Figure 7 presents the output prediction and target labels of seven classes. Each class represents the combination of three occupants' presence state. The top figure shows the prediction using k -NN^{seq}, while the bottom figure shows the prediction using sequence RNN. Using k -NN^{seq} from 08.30 until 09.00 AM has several miss-classified class-5 due to the power consumption raising to 80W. RNN could deal with this power consumption better until 12.15 AM. While around 02.00 PM both classifiers failed in detecting class-7. The classifier based on RNN performed better in recognizing class-3 at 02.15 PM. In the same period, the k -NN^{seq} classifier mostly miss-classified as a class-6 until 03.00 PM. Figure 8 presents the people count estimation given the occupancy prediction. The sequence-based classification using RNN provides fewer spikes and estimates occupancy level more precisely than k -NN^{seq}.

V. DISCUSSION

The present work describes three different approaches to detect occupancy using various machine learning classifiers. We illustrated the performance of the classifiers and its fluctuation depending on how the data is being processed; namely, whether or not it is shuffled prior to the training and testing phase, and whether or not we take its temporal ordering in account.

A. INDIVIDUAL BEHAVIOUR OBSERVATION

During the two-month power consumption observation, we noticed that the participants had a strong tendency to start or stop work at a particular time of the day, as shown in Figure 5. For example, P1 frequently left at 6.00 PM, P2 arrived at 10.00 AM for more than half of their work days and two-thirds of P3's work days consistently started at around 10.00 AM. This pattern might help classifiers to recognize people, if temporal features were introduced.

In our experimental setup, we used two Smappee power meters to collect power usage data. This setup is only realistic in an experimental setting, and would be over-optimistic for a real world environment, in which at most one power meter would be available.

The total amount of consumed electricity strongly depends on the used appliances. The more heterogeneous the appliances frequently being used, the higher the variance in the power consumption distribution. As shown in Figure 6, P3's data show double notable peaks (two notably different power profiles). It could be attributed to the fact that P3 uses different devices in his work sessions (e.g., a PC, laptop, or a combination of them).

In Table 2, we validate that the actual presence of P1 and P3 at their workspace were accurately detected using the threshold approach on the individual power consumption. It is because P1 and P3 consistently used office-related appliances during their presence, so that occupancy can be inferred based on threshold-based ground truth. This finding supports the generated labels to be used as the ground truth complement. In contrast to other people, however, the predictions of P4's occupancy were poor. This result may be explained by the fact that P4 did not always use PC during the observation period. The low value of recall shows that the correct prediction of P4 being present (true positive) was quite small compared to the real presence. That is, P4 was frequently present in the workspace when his power consumption below the threshold. It happened when P4 read literature during the week and letting the PC going idle or to sleep. It does not change the legitimacy of the generated labels as ground truth since we believe P2 has the same work patterns (i.e., constantly



FIGURE 7. Occupancy detection with ID on the 10th September 2018 using k -NN with a kappa of .717 (top) and RNN with a kappa of .836 (bottom). Blue line shows active power in Watts, orange line shows labels of predicted occupancy, and dashed green line shows labels of real occupancy.

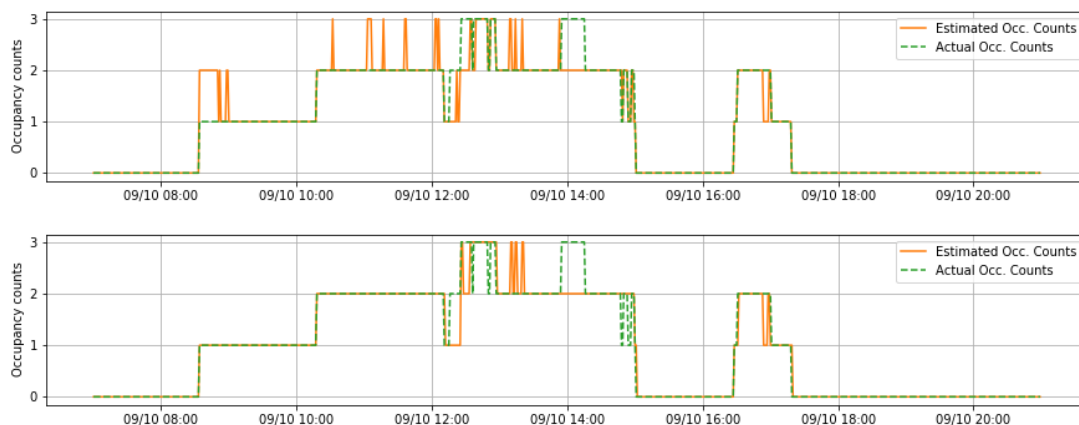


FIGURE 8. Occupant counting on the 10th September 2018 using k -NN with a kappa of .859 (top) and RNN with a kappa of .888 (bottom).

working with a computer), while P4 was not involved in the dataset used in machine-learning based occupancy detection.

B. OCCUPANCY DETECTION

The advantage of the sequential data classification and having more features can be seen in Table 3 and Table 4, respectively (both in the training and validation sets). We can see that RNN achieved the best result over 2-minute sequences. RNN can use the temporal ordering by using earlier observations to make predictions. However, this benefit is not available for the adapted version of k -NN for sequence classification

as the best result is obtained using the 1-minute sequence due to k -NN nature. It is easier and more accurate to find similar instance rather finding similar sequence in provided training set. For both approaches, adding all features obtained by the meter and the time of day improves the classification performance.

RNN achieves an approximately 5% higher kappa measure than the regular neural network approach, but there is only less than 1% improvement in the sequential analysis using the adapted k -NN as shown in Table 6. This finding suggests that an improvement can be achieved by looping through

the information from the previous input values to predict outcomes of the following instances in a sequence. As for the k -NN, the nearest neighbor based algorithm works based on the majority voting labels of the nearest samples to the query. Hence, the performance solely relies on sample availability, regardless the sequential ordering. The final evaluation on the test shows that the results of the k -NN^{seq} and RNN algorithms are comparable, with accuracy between .96 and .97, a kappa between .93 and .94, and F-measure around .91.

Comparing with previous works, this paper improves the classification performance while significantly reducing costs and level of intrusiveness, as it uses a single power meter. In fact, Akbar *et al.* deployed one meter per user to determine occupancy states with .80 – .94 F-measure [7]. In that work, four meters were used to detect the presences of four individuals. Zhao *et al.* used one meter per device to classify computer activation states and employees' occupancy [11]. Twenty eight power meters were used to detect 10 employees. The authors reported that .90 accuracy and .69 Kappa could be reached using their proposed approach. Shetty *et al.* reported .94 averaged accuracy of three people, but they failed to detect the absence states of another person [10].

Among the evaluated techniques, FHMM performs the worst. Its weak performance could be explained by the fact that this approach finds the most probable occupancy states given the observed power consumption, instead of finding separation boundaries among different occupancy states. Moreover, we only consider two individual states (i.e., being present or absent). There is no definition of the other states that might lead to different power consumption patterns (e.g., being present using only one of two available monitors without charging laptop). The model simplification (i.e., by providing only active power as a feature) might also negatively influence the results.

C. DAILY OCCUPANCY DETECTION

The aim of daily occupancy detection was to see the classification performance in the most recently collected test set. This scenario is to represent a chance in daily life condition where measurements are fresh and not represented in any cross-validation folds.

In this scenario, we find lower kappa measures in classification, as shown in Table 7, compared to the results from the stratified random test set that reach a kappa measure of .93 to .94 in Table 6. This might be due to the variance of power consumption that one might find in practice. As we only provided the first 85% portion of the data, we did not introduce samples of the whole data to the classifiers (see Section III-B1). In this particular case, it is apparent that RNN generally performs better in the most recently collected test set (e.g., in the five of seven work days, as shown in Table 7) compared to k -NN^{seq}. This finding can be seen in Figure 7 during a period from 08.30-09.00 AM and 02.15-03.00 PM. In the former, the power consumption fluctuates between 80 and 50 Watts. k -NN^{seq} misclassifies class 1 as class 5 when the power consumption reaches 80 Watts, while RNN correctly

infers class-1. In the latter, k -NN^{seq} misclassifies class 3 as class 6 while RNN correctly infers class 3. A possible explanation for this might be that RNN regards the output of the previous instances to predict the output of the current instance which is completely unseen in training. In contrary, k -NN^{seq} easily miss-classifies if there are very similar samples that belong to different classes. k -NN^{seq} outperforms RNN when it finds many similar samples with the same label previously presented in training set.

VI. CONCLUSION AND FUTURE WORK

The present work investigates low-intrusive power-metering for occupancy detection in a shared office environment. We proposed a setup and method for such detection and evaluated the approach experimentally. In our experiment, we measured the aggregated power consumption of three occupants during a two-month period, and found that individual occupancy detection with user identification is feasible with a kappa measure reaching .93. The contribution of this work lies in: (i) An approach using aggregate power consumption for fine grained occupancy detection, and (ii) an evaluation of classification techniques for the specific problem; namely, the use of sequential classification with RNN, for unseen instances during the training phase.

Initial experimental data about occupant behavior was done via power readings in order to obtain a general understanding of typical occupancy in the office. In our case, the participants came to the office without a strict schedule. Nevertheless, we found some regularities in the data. Some people regularly came to or left from the office at a particular time of the day. We further observed that two of three participants are highly likely to work using a PC.

Based on empirical evaluation, occupancy detection of individuals with a fine granularity level (i.e., either distinguishing one person among the others or counting occupants) is feasible for three people in a shared office. We achieved the best performance using k -NN and recurrent neural network with a kappa measure of about .93 in the stratified random test set.

We tested with the non-shuffled data set in order to provide an overview of daily occupancy (i.e. when testing data has never been introduced in the training set. In this specific case, the RNN LSTM seems to outperform k -NN in the five of seven workdays. The performance ranged from .559% to .877% and between .628% and .952% with k -NN and RNN, respectively. If we are interested only in occupant counting (i.e., ignoring which people are present), the performance is higher, as it disregards incorrect identification and focuses only on the estimation number of people.

Our results are promising, especially when compared to other approaches for occupancy detection. In terms of occupancy granularity, our approach has a higher level of detail than PIR sensors based ones [25]–[27], [38]. While the accuracy is not as high as with a Kinect camera [39], power meter-based approach is cheaper and less intrusive in terms of people monitoring. In fact, avoiding the use of cameras and

instead monitoring power consumption helps people feel less observed and controlled.

Compared with related work, that uses 3-9 power meters [7], [10], [12] or even 28 meters [11] for detecting three people in an office, we achieve similar precision with just one power meter. Our proposed system currently has a cost of about 250 EUR which is 29.6% cheaper than the one proposed by Shetty et al. [10] and 16% cheaper than that proposed by Akbar et al. [7]. BLE based occupancy detection is incomparable to our approach as this modality classifies a location of a person rather than classifying occupants of a room. Still, a combination of BLE with our approach might be an interesting virtue for future work.

Some limitation exists in the present work. First, the result is based on a specific experimental setup. A similar experiment with different setups, rooms, climates, and subjects might show other values. For example, involving more occupants in an office with homogenous appliances might make more challenging the distinction of occupants presences. Furthermore, data collection and retraining models need to occur whenever there is a change in subjects occupying regularly a room (e.g., joining in or moving out from the office). This limitation might be resolved by guiding the occupants in collecting training data by themselves using an application. The proposed approach, however, might not work in an office where individual desks are not assigned to users. Secondly, this study did not evaluate the scalability of the approach. While the proposed solution works well with three users attached to a single power meter, we did not evaluate the performance of the classification when more people use appliances connected to the same power meter. Evaluating the scalability of our approach is an interesting direction for future work. A third limitation relates to the implementability of the present approach with respect to privacy. Although the proposed approach only takes power meter data (instead of images or video recordings, or other intrusive systems), our solution might impact privacy. Fourth, the best algorithms in this work are supervised learning approaches, meaning that labeled training data must be available in order to build models. Hence, the approach is limited to training data availability. Finally, our system cannot distinguish between people with a similar power consumption, or without any consumption. As such, we believe our approach requires a fusion with a different source of information (e.g., BLE beacons). A further study with a focus on building such a combined classifier is therefore suggested.

REFERENCES

- [1] S. Naylor, M. Gillott, and T. Lau, "A review of occupant-centric building control strategies to reduce building energy use," *Renew. Sustain. Energy Rev.*, vol. 96, pp. 1–10, Nov. 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032118305264>
- [2] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, and Y. Agarwal, "Sentinel: Occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings," in *Proc. 11th ACM Conf. Embedded Netw. Sensor Syst. (SenSys)*, New York, NY, USA, 2013, pp. 17:1–17:14, doi: [10.1145/2517351.2517370](https://doi.org/10.1145/2517351.2517370).
- [3] F. Nizamic, T. A. Nguyen, A. Lazovik, and M. Aiello, "GreenMind—An architecture and realization for energy smart buildings," in *Proc. Conf. ICT Sustain. (ICT4S)*, Amsterdam, The Netherlands: Atlantis Press, Aug. 2014, pp. 1–10, doi: [10.2991/ict4s-14.2014.3](https://doi.org/10.2991/ict4s-14.2014.3).
- [4] W. Shen, G. Newsham, and B. Gunay, "Leveraging existing occupancy-related data for optimal control of commercial office buildings: A review," *Adv. Eng. Inform.*, vol. 33, pp. 230–242, Aug. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1474034616301987>
- [5] A. Ebadat, G. Bottegai, D. Varagnolo, B. Wahlberg, and K. H. Johansson, "Regularized deconvolution-based approaches for estimating room occupancies," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 4, pp. 1157–1168, Oct. 2015.
- [6] T. Petrovic, K. Echigo, and H. Morikawa, "Detecting presence from a WiFi router's electric power consumption by machine learning," *IEEE Access*, vol. 6, pp. 9679–9689, 2018.
- [7] A. Akbar, M. Nati, F. Carrez, and K. Moessner, "Contextual occupancy detection for smart office by pattern recognition of electricity consumption data," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 561–566.
- [8] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, "Occupancy detection from electricity consumption data," in *Proc. 5th ACM Workshop Embedded Syst. Energy-Efficient Buildings (BuildSys)*, New York, NY, USA, Nov. 2013, pp. 10:1–10:8, doi: [10.1145/2528282.2528295](https://doi.org/10.1145/2528282.2528295).
- [9] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, and P. Shenoy, "Non-intrusive occupancy monitoring using smart meters," in *Proc. 5th ACM Workshop Embedded Syst. Energy-Efficient Buildings (BuildSys)*, New York, NY, USA, Nov. 2013, pp. 9:1–9:8, doi: [10.1145/2528282.2528294](https://doi.org/10.1145/2528282.2528294).
- [10] S. S. Shetty, H. D. Chinh, M. Gupta, and S. K. Panda, "User presence estimation in multi-occupancy rooms using plug-load meters and PIR sensors," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.
- [11] J. Zhao, B. Lasternas, K. P. Lam, R. Yun, and V. Loftness, "Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining," *Energy Buildings*, vol. 82, pp. 341–355, Oct. 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378778814005714>
- [12] M. Jin, R. Jia, Z. Kang, I. C. Konstantakopoulos, and C. J. Spanos, "PresenceSense: Zero-training algorithm for individual presence detection based on power monitoring," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Buildings (BuildSys)*, New York, NY, USA, Nov. 2014, pp. 1–10, doi: [10.1145/2674061.2674073](https://doi.org/10.1145/2674061.2674073).
- [13] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, "A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring," *Appl. Energy*, vol. 185, pp. 331–344, Jan. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S030626191631488X>
- [14] K. Basu, V. Debusschere, A. Douzal-Chouakria, and S. Bacha, "Time series distance-based methods for non-intrusive load monitoring in residential buildings," *Energy Buildings*, vol. 96, pp. 109–117, Jun. 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0378778815002133>
- [15] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.
- [16] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Bajic, "AMPds: A public dataset for load disaggregation and eco-feedback research," in *Proc. IEEE Electr. Power Energy Conf.*, Aug. 2013, pp. 1–6.
- [17] A. G. Ruzzelli, C. Nicolas, A. Schoofs, and G. M. P. O'Hare, "Real-time recognition and profiling of appliances through a single electricity sensor," in *Proc. 7th Annu. IEEE Commun. Soc. Conf. Sensor Mesh Ad Hoc Commun. Netw. (SECON)*, Jun. 2010, pp. 1–9.
- [18] A. S. N. U. Nambi, A. R. Lua, and V. R. Prasad, "LocED: Location-aware energy disaggregation framework," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ. (BuildSys)*, 2015, pp. 45–54. [Online]. Available: <http://www.es.wi.tudelft.nl/papers/2015-Uttama-LockED.pdf>
- [19] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ. (BuildSys)*, New York, NY, USA, Nov. 2015, pp. 55–64, doi: [10.1145/2821650.2821672](https://doi.org/10.1145/2821650.2821672).
- [20] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Sci. Data*, vol. 2, no. 1, Dec. 2015, Art. no. 150007.
- [21] A. Zoha, A. Gluhak, M. Nati, and M. A. Imran, "Low-power appliance monitoring using factorial hidden Markov models," in *Proc. IEEE 8th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process.*, Apr. 2013, pp. 527–532.

- [22] A. R. Pratama, F. J. Simanjuntak, A. Lazovik, and M. Aiello, "Low-power appliance recognition using recurrent neural networks," in *Applications of Intelligent Systems*, vol. 310. Amsterdam, The Netherlands: IOS Press, 2018, pp. 239–250.
- [23] A. R. Pratama, Widyawan, A. Lazovik, and M. Aiello, "Power-based device recognition for occupancy detection," in *Service-Oriented Computing—ICSOC 2017 Workshops*, L. Braubach, J. M. Murillo, N. Kaviani, M. Lama, L. Burgueño, N. Moha, and M. Oriol, Eds. Cham, Switzerland: Springer, 2018, pp. 174–187.
- [24] Z. Kang, M. Jin, and C. J. Spanos, "Modeling of end-use energy profile: An appliance-data-driven stochastic approach," in *Proc. 40th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2014, pp. 5382–5388.
- [25] Y. Zhao, W. Zeiler, G. Boxem, and T. Labeodan, "Virtual occupancy sensors for real-time occupancy information in buildings," *Building Environ.*, vol. 93, pp. 9–20, Nov. 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360132315300330>
- [26] T. Ekwuegbe, N. Brown, V. Pakka, and D. Fan, "Improved occupancy monitoring in non-domestic buildings," *Sustain. Cities Soc.*, vol. 30, pp. 97–107, Apr. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2210670716303808>
- [27] A. Arora, M. Amayri, V. Badarla, S. Ploix, and S. Bandyopadhyay, "Occupancy estimation using non intrusive sensors in energy efficient buildings," in *Proc. 14th Conf. Int. Building Perform. Simulation Assoc.*, Hyderabad, India, Dec. 2015, pp. 1–9. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01864941>
- [28] B. Zhang, Y. Shen, and L. Zhang, "Inferring building occupancy based on statistical modeling of multi-sensor data," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2016, pp. 809–814.
- [29] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer Series in Statistics). New York, NY, USA: Springer, 2001.
- [30] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Berlin, Germany: Springer-Verlag, 2006.
- [31] G. Shakhnarovich, T. Darrell, and P. Indyk, *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice* (Neural Information Processing). Cambridge, MA, USA: MIT Press, 2006.
- [32] H. B. Demuth, M. H. Beale, O. De Jess, and M. T. Hagan, *Neural Network Design*, 2nd ed. Boulder, CO, USA: Martin Hagan, 2014.
- [33] F. J. Simanjuntak. (Aug. 2017). *Deep Learning Approach for Electric Appliances Recognition*. [Online]. Available: <http://ugm.id/fsimanjuntak>
- [34] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," in *Proc. Adv. Neural Inf. Process. Syst.*, 1996, pp. 472–478.
- [35] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [36] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960, doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104).
- [37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [38] P. Liu, S.-K. Nguang, and A. Partridge, "Occupancy inference using pyroelectric infrared sensors through hidden Markov models," *IEEE Sensors J.*, vol. 16, no. 4, pp. 1062–1068, Feb. 2016.
- [39] S. Petersen, T. H. Pedersen, K. U. Nielsen, and M. D. Knudsen, "Establishing an image-based ground truth for validation of sensor data-based room occupancy detection," *Energy Buildings*, vol. 130, pp. 787–793, Oct. 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S037877881630809X>



AZKARIO RIZKY PRATAMA received the bachelor's and master's degrees from the Universitas Gadjah Mada, Indonesia, and the Ph.D. degree in computer science from the University of Groningen, The Netherlands. He is currently affiliated as a Lecturer with Universitas Gadjah Mada. His research interests include context-awareness, pervasive and mobile computing, occupancy detection, and electric load identification.



FRANK JOHAN BLAAUW received the Ph.D. degree in computer science from the University of Groningen, in 2017. He founded Researchable, a company that builds software and focuses on scientific software and data driven innovations, in 2018. He is specialized in software engineering, distributed systems, and machine learning.



ALEXANDER LAZOVIK received the Ph.D. degree in computer science from the University of Trento, in 2006. He is currently a Professor of computer science with the Distributed Systems Group, University of Groningen, after being an ERCIM Fellow at CWI and INRIA, and an Intern at IBM T. J. Watson. His main research interests include AI planning and discrete optimization in highly distributed environments, having the IoT as a main application domain. His research results

have been field-tested in a number of existing buildings serving as living labs.



MARCO AIELLO (Senior Member, IEEE) received the Ph.D. degree in logic from the University of Amsterdam, the Habilitation degree in applied informatics from TU Wien, and the master's degree in engineering from the La Sapienza University of Rome. He is currently a Professor of computer science and the Head of the Service Computing Department, University of Stuttgart, Germany. An Elected Member of the European Academy of Sciences and Arts, he is an Honorary Professor of distributed systems with the University of Groningen, The Netherlands. His research interests include smart energy systems, smart buildings, service computing, and spatial reasoning.

...