Virtual Occupancy Sensing:
Using Smart Meters to Indicate Your Presence

Ming Jin, *Member, IEEE,* Ruoxi Jia, *Member, IEEE,* and Costas J. Spanos, *Fellow, IEEE*

Abstract—Occupancy detection for buildings is crucial to improving energy efficiency, user comfort, and space utility. However, existing methods require dedicated system setup, continuous calibration, and frequent maintenance. With the instrumentation of electricity meters in millions of homes and offices, however, power measurement presents a unique opportunity for a non-intrusive and cost-effective way to detect occupant presence. This study develops solutions to the problems when no data or limited data is available for training, as motivated by difficulties in ground truth collection. Experimental evaluations on data from both residential and commercial buildings indicate that the proposed methods for binary occupancy detection are nearly as accurate as models learned with sufficient data, with accuracies of approximately 78 to 93% for residences and 90% for offices. This study shows that power usage contains valuable and sensitive user information, demonstrating a virtual occupancy sensing approach with minimal system calibration and setup.

Index Terms—Occupancy detection, electricity, non-intrusive monitoring, energy-efficient buildings

1 Introduction

Technological innovations in sensor networks and mobile computing extend the frontiers of building science and opens up a wealth of opportunities. A combination of factors contribute to the connectedness of pervasive sensing nodes: lowered manufacturing cost, miniaturized size and enhanced functionality, integrated sensing services, visualization, and actuation, and raised awareness of energy, privacy, and security issues.

Individuals spend 90% of their time in buildings, either in their homes or in their offices. Buildings consume approximately 40% of all energy usage in the United States. Heating, ventilation, and air conditioning (HVAC) and lighting are responsible for 43% and 11% of energy usage in residential buildings, and 33% and 26% of energy usage in commercial buildings [1]. Among various measures for energy savings, such as building renovations, dynamic pricing, and social games, occupancy sensing has been shown to contribute to both occupant comfort and energy efficiency [2], [3]. The current study focuses on binary-level presence detection (i.e., $0/1$ decision – a room or a house is either occupied or vacant), as opposed to the occupancy of an identified individual or the number of people in a region. Binary information is usually sufficient for applications like automated zonal control for indoor temperature, air quality, and lighting [2], [3], [4], [5], as well as health monitoring in residential buildings [6].

Existing approaches to occupancy detection use either reliable but specialized sensors, such as cameras [3], passive infrared (PIR) [7], magnetic reed switches [2], [3], active RFID badges [5], or non-specialized for occupancy detection yet fallible environmental measurements, including CO_2 [8], [9], temperature [10], and particulate matters [11]. With the installation of electricity meters in millions of homes and offices [12], which are continuously monitoring power usage, one question comes to mind:

"Can energy consumption be used to indicate one's presence?"

While previous work that relies on power measurements focuses on situations where sufficient data is available from target households [10], [13], [14], [15], [16], [17], collecting the occupancy ground truth that is required to train the detection algorithms is a challenging process. It often relies on 1) user engagement, e.g., completing a survey every 15 or 60 minutes [13], [16], or uploading the GPS trajectory data as a proxy for their presence ground truth [14], 2) instrumentation efforts, e.g., setting up a wireless network of cameras, motion sensors, or door switches [10], [17]. A systematic approach that relaxes the requirement of ground truth data collection is desirable for practical application.

The following scenarios illustrate some emerging applications of smart meters [18], in which user presence information can bring direct benefits:

Scenario 1 (S1): Occupants in an office building participate in a social game to save energy by having their cubicles instrumented with power meters [19]. Building managers want to correlate the energy consumption pattern of the building with the schedules of its occupants and assess its energy efficiency. However, they prefer an easy and non-intrusive way to access the occupant presence information without additional infrastructures or surveys.

Scenario 2 (S2): A start-up company on smart meters for home automation has collected a substantial amount of occupancy and electricity data from incentivized households to train its presence detection system. A targeted product feature is effortless calibration by requiring minimal inputs from new customers while maintaining comparable performances.

• *M. Jin, R. Jia, and C. J. Spanos are with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, 94720, United States*
E-mails: {jinming, ruoxijia, spanos}@eecs.berkeley.edu

Manuscript received ; revised

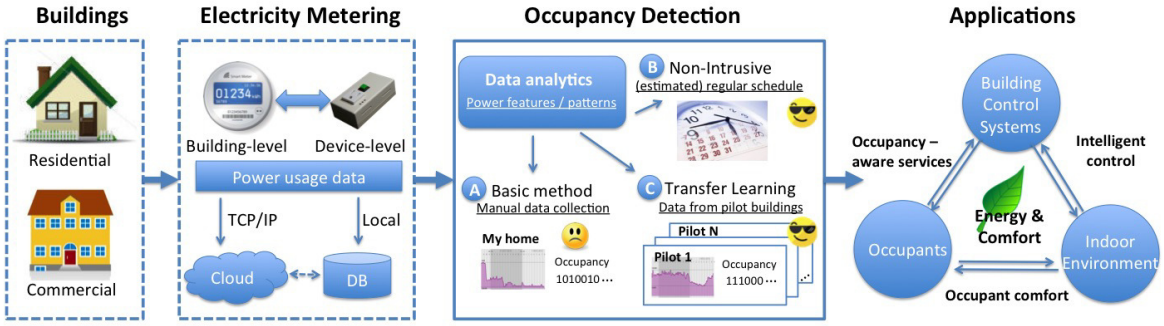


Fig. 1: Key system components that offer occupancy-aware services that are enabled by presence detection.

In both situations, smart meters are viable proxies for demand approximation [18] (S1) and/or home automation [20] (S2) by extending basic metering functionality to presence detection. The benefits of higher detection accuracy include more substantial energy savings and reliable automation systems with no additional costs to instrument sensors [2]. Nevertheless, these two problems differ in training source availability: While S1 is characterized by very limited presence data accessible to the learner, while S2 features extensive data but not from its customers' households.

Hence, the objective of this study is to develop solutions to address the issues that arise while learning from limited or no training data. Key components of the end system include 1) electricity measurements from building- and/or device- level meters, 2) occupancy detection based on power usage, e.g., Non-intrusive Learning (NL) or Transfer-Learning (TL) methods proposed in the study, and 3) occupancy-aware applications, e.g., demand controlled ventilation [2], [5], [9], [21] and lighting [19], and building-level demand response [22] (Fig. 1).

Our main contributions are summarized below:

- Formulates the problems as Base Learning (BL), Non-intrusive Learning (NL), and Transfer Learning (TL).
- Explores power features and studies the window and sampling effects as a practice guide.
- Develops and evaluates solutions to NL and TL for residential and commercial buildings, which achieves around 80 to 90% accuracy.

The rest of the paper is organized as follows. Formulations of BL, NL, and TL are described in § 2. § 3 introduces the methods, in particular, multiview-based iteration and surrogate loss methods for NL (§ 3.2), and the transfer learning optimization method for TL (§ 3.3), as well as explores power features (§ 3.4). BL, NL, and TL datasets and results are reported and discussed in § 4. While related work is summarized in § 5, § 6 concludes this study and proposes future work.

2 Problem Formulations

In this section, we first describe the basic framework, and then formulate the problems that prompted the solutions proposed in the study.

2.1 Basic Framework

We set the basic framework of our learning problem, in which the notations follow the literature from statistical decision theory. We use "presence" and "occupancy" interchangeably in the texts.

The dataset, $S = \{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_n, y_n)\}$, includes electricity and occupancy data. The presence information, $y \in \mathcal{Y} = \{-1, +1\}$, takes either the value of $+1$ or -1, indicating the state of presence (occupancy) or absence (vacancy), which are often provided by users [15] or reliable sensors like camera [3], ultrasonic [23], and passive infrared (PIR) sensors [7]. The measurement data, $\boldsymbol{x} \in \mathcal{X} = \times_{j=1}^{m} X_j$, consists of m distinct features derived from electricity and time, as described in § 3.4. We make the following assumptions in the design of occupancy detectors.

Assumption 1: While an individual's energy consumption and occupancy behaviors are stable over the period of interests, two individuals' consumption and occupancy behaviors may differ.

Assumption 2: Samples in S are independent and identically distributed (i.i.d.).[1]

Our goal is to learn the presence detector f that minimizes the risk, $R_{l,\mathcal{F}}(f) = \mathbb{E}_{f \in \mathcal{F}}\left[l(y, f(\boldsymbol{x}))\right]$, where $\mathcal{F}$ delineates the range of detectors, for instance, the class of linear classifiers, and $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is the loss function, which penalizes mistakes in detection. An example is the 0-1 loss $l_{01}(y, f(\boldsymbol{x})) = \mathbf{1}(y \neq f(\boldsymbol{x}))$ that results in 1 if the two arguments differ and 0 otherwise.

Since we can only access the unknown power consumption behaviors through the dataset S, our strategy is to search for the *empirical risk minimizer (ERM)* as given by:

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} l(f(\boldsymbol{x}_i), y_i) \tag{1}$$

The *regret* characterizes how "poorly" the learner has performed compared to the performance of the best achievable learner in the class, as given by

$$\delta R_{l,\mathcal{F}}(f) = \underbrace{R_{l,\mathcal{F}}(f)}_{\text{risk of } f} - \underbrace{\inf_{f' \in \mathcal{F}} R_{l,\mathcal{F}}(f')}_{\text{risk of the best detector}} \tag{2}$$

where a tight upper bound on the regret suggests a performance guarantee of the algorithm when used in practice.

1. Although time series data may have time auto-correlation issues, they are resolved when the data is organized into groups of 10 to 15 minutes, thereby justifying the i.i.d. assumption.

2.2 Problem Formulations

Presence learning problems can be described as follows:

- **Base Learning (BL):** Find the optimal detector given the power and occupancy data from a user, $\mathcal{S}$.
- **Non-intrusive Learning (NL):** Find the corresponding occupancy classifier with only the electricity usage data.
- **Transfer Learning (TL):** Learn a presence detector for the target home where limited data is available based on sufficient power and occupancy data from a group of similar households.

The set up of BL, as in most previous studies [10], [14], [15], [16], [17], [24], requires the collection of presence and energy data for an extended period of time to train the classifier. Then, the classifier makes presence decisions when new electricity measurement information is provided.

NL, in comparison, corresponds to Scenario 1. Instead of intrusively collecting user's states for training purposes, the learning process is based on electricity data alone.

TL refers to Scenario 2, in which several individuals (households), or source, have been surveyed for a prolonged period of time. But, only limited data has been obtained for the individuals of interests, or target. The objective is to transfer the knowledge from the source to the target, and improve the detection accuracy of the target task.

2.3 Discussion

Due to the difficulties in collecting the occupancy ground truth data, there is increasing demand to move away from BL and shift toward NL and TL.

The design of occupancy detection algorithm involves trade-offs among the following practical concerns:

Accuracy. Generally, a higher rate of accuracy for the algorithm is preferred. But, the cost of an erroneous detection differs based on the presence system's application and issue severity. For instance, if a company's alarm system fails to detect an intrusion, the issue severity and resulting cost are high. However, if a home automation system falsely assumes presence, the issue severity and resulting cost are low [25].

Costs. The *capital costs* are due to the smart meters quantity and complexity, as well as the communication and storage infrastructures. The *training costs* involve the time and labor invested in collecting presence data for system calibration, which might be different for recruited individuals and new customers.

The solution to the more challenging NL and TL problems is based on three key observations:

1) Energy consumption differs markedly when individuals are present versus when they are absent.
2) Office buildings are usually vacant during non-business hours.
3) As noted by Nobel laureate Kahneman in his book "Thinking, Fast and Slow" [26], people's behaviors are patterned (the "slow" system) with spontaneous deviations (the "fast" system). These patterns are consistent over time.

While the third point lays the foundation for tackling the TL problem, the first two points lead to the solutions to the NL problem, as detailed in the next section.

3 Methodology

The core of the presence detection methods is embodied in the answers to these two questions: 1) What is the set of features derived from power usage that is most indicative of user presence, and 2) How can one estimate this correspondence. BL approaches approximate this relationship based on training data from the target. Due to a lack of direct access to data for NL and TL, the "multiview" and "transferability" assumptions have been created. They are detailed in § 3.2 and § 3.3, respectively. § 3.4 illustrates the design and selection of power features that address the first question about the features that are the most indicative of user presence.

3.1 Base Learning Methods

The BL goal is to find the learner $\hat{f}$ that can minimize the empirical loss (1). A variety of models have been proposed, such as naïve Bayes, Bayes net, logistic regression, support vector machines (SVM), adaptive boosting (AdaBoost), decision tree (J48), and random forest. These models follow the same ERM framework but differ in the loss function selection.

For instance, SVM optimizes the hinge loss: $l_{hinge}(f(\boldsymbol{x}_i), y_i) = \max(0, 1 - y_i f(\boldsymbol{x}_i))$ where $f(\boldsymbol{x}_i) = \langle \boldsymbol{\omega}, \boldsymbol{x}_i \rangle$ is linear and parameterized by $\boldsymbol{\omega} \in \mathbb{R}^m$. Intuitively, it seeks a decision boundary that maximizes the distance from the two classes, with a special focus on those near the boundary, a.k.a., the "support vectors".

The loss function reflects how the model treats the misclassification error, which, together with the function that predicts each instance, offers a summary of the method. More details about the BL algorithms can be found in [27]. BL algorithm implementations are also available, for example, LibSVM [28] and Weka [29]. Experimental results are reported in § 4.2.

As BL's framework generally requires substantial training data, it is not suitable if user convenience is key. In the next section, we discuss how the NL and TL solutions satisfy the user convenience requirement by modifying the loss function or optimization problem, e.g., surrogate loss in § 3.2.2, and knowledge transfer in § 3.3.

3.2 Non-intrusive Learning Methods

Unlike BL, the solution to the NL problem requires the use of problem-specific structures as we lack sufficient data to estimate the presence – electricity correspondence. The key is to use time as clues to infer occupant presence. As individuals are habitual, this demonstrates consistency principle for commercial and residential buildings (Fig. 2).

Let $\mathcal{X}_{v1}$ and $\mathcal{X}_{v2}$ be two "views" that partition the original feature space, $\mathcal{X} = \mathcal{X}_{v1} \times \mathcal{X}_{v2} = \times_{j=1}^{m} X_j$, and $\mathcal{F}_{v\cdot}$ be the associated function space. Our methods are based on the following assumption:

Multiview Assumption: The risks $R_{l,\mathcal{F}_{v\cdot}}(f^*_{v\cdot})$ of learners trained by each "view" $v\cdot$ separately, $f^*_{v\cdot} = \arg\min_{f \in \mathcal{F}_{v\cdot}} \mathbb{E}\left[l(y, f(\boldsymbol{x} \in \mathcal{X}_{v\cdot}))\right]$, is comparable to the risk, $\mathbb{R}_{l,\mathcal{F}}(f^*)$, of the learner trained on the complete feature space, $f^* = \arg\min_{f \in \mathcal{F}} \mathbb{E}\left[l(y, f(\boldsymbol{x} \in \mathcal{X}))\right]$.

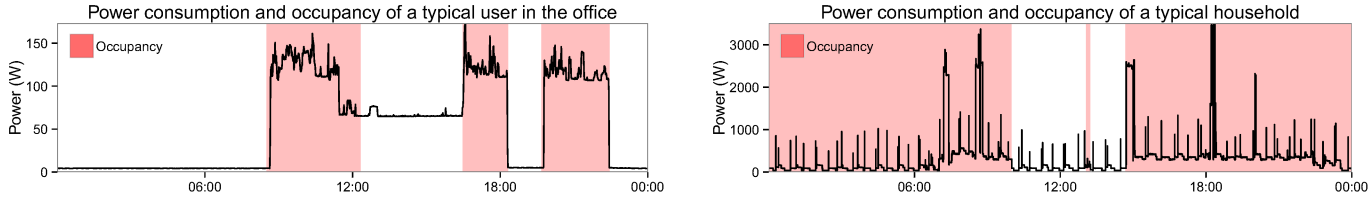


Fig. 2: Typical daily power consumption of an occupant in a commercial building (left) and in a household (right). The red color indicates user presence while the blue color indicates user absence.

The multiview assumption states that each view, power or time, is sufficient for the problem at hand, as suggested by Blum and Mitchell [30].

Next, we describe two approaches, multiview-based iteration training (MIT) and surrogate loss. Based on our previous work [23], the first method initializes the training set by the time view, and iteratively refines the labels until a stopping criterion is met. Built on the work of Natarajan et al. [31], the second method uses the surrogate loss in replacement for the original loss, in order to "reverse" the effect of labels' corruption introduced to the training data by the time view.

3.2.1 Multiview-based Iteration Training

Assign the first view, $\mathcal{X}_{v1}$, as the time of the day, and the second view, $\mathcal{X}_{v2}$, as the power-derived features, and let f_{v1} and f_{v2} denote the classifiers based on the two views respectively. Given the unlabeled dataset $S = \{(\boldsymbol{x}_1^{v1}, \boldsymbol{x}_1^{v2}), \cdots, (\boldsymbol{x}_n^{v1}, \boldsymbol{x}_n^{v2})\}$, where $\boldsymbol{x}_i^{v1} \in \mathcal{X}_{v1}$ and $\boldsymbol{x}_i^{v2} \in \mathcal{X}_{v2}$, MIT proceeds as follows (the pseudo-code can be found in the appendix):

1) *Initialization:* Set the initial training set according to the prior information provided by the time view, or (rough) occupancy schedules.
2) *Multiview training:* For rounds $t = 1, 2, ...,$ train the power-based classifiers, and determine the new guesses based on the majority votes among the power- and time-based classifiers.
3) *Labeled set updates:* Perform the following update:

$$L_j^{t+1} = \{L_j^t \cap L_{j,n}^t\} \cup Sample\{L_j^t \Delta L_{j,n}^t; \alpha_j\} \quad (3)$$

for $j \in \{-1, +1\}$, where $L_j^t = \{((\boldsymbol{x}_i^{v1}, \boldsymbol{x}_i^{v2}, \hat{y}_i) | \hat{y}_i = j\}$, and $L_{j,n}^t$ denote the set of samples whose labels are newly classified as j, $L_j^t \Delta L_{j,n}^t$ is the symmetric difference set operation, and α_j is the sampling rate for label $j \in \{-1, +1\}$.
4) *Stopping condition:* Stop the iteration whenever (7) in Theorem 3.3 is satisfied.

The dataset update (3) keeps the samples in which the labels are agreed upon in the two successive iterations, and insert a random subset of "controversial" samples, with $\alpha_j \in [0, 1]$ controlling the learning rate. The stopping rule (7) is triggered when it seems unlikely that additional iterations can improve the accuracy rate.

Different than the BL, MIT operates in a condition such that the noise in the training set can not be ignored. The probably approximately correct (PAC) framework proposed by Valiant [32] can be applied to resolve this issue.

Theorem 3.1. *[32] If we draw a sequence of*

$$n \geq \frac{2}{\epsilon^2 (1 - 2\eta)^2} \log\left(\frac{2N}{\delta}\right) \quad (4)$$

samples from a distribution and find any hypothesis f_i that minimizes disagreement with the training labels, where ϵ is the hypothesis worst-case classification error rate, η is the upper bound on the training noise rate, N is the number of hypotheses, and δ is the confidence, then the following PAC property is satisfied:

$$\mathbb{P}\left[d\left(f_i, f^*\right) \geq \epsilon\right] \leq \delta \quad (5)$$

where $d(,)$ is the sum over the probability of elements from the symmetric difference between the two hypotheses f_i and the optimal f^*.

The above theorem provides a high probability bound on the classification error given the training noise rate, which can be estimated as follows. If we let $L_{-1}^t = L_{-1,\checkmark}^t \cup L_{-1,\times}^t$, where $L_{-1,\checkmark}^t$ and $L_{-1,\times}^t$ are the set of true negative and false positive samples, respectively, similarly $L_{+1}^t = L_{+1,\checkmark}^t \cup L_{+1,\times}^t$, also $U^t = U_{-1}^t \cup U_{+1}^t$ is the set of unlabeled samples, then the training noise rate, η_t, is given by:

$$\eta_t = \frac{|L_{-1,\times}^t| + |L_{+1,\times}^t|}{|L_{-1}^t| + |L_{+1}^t|} \quad (6)$$

Assume the hypothesis makes classification errors at the rate ϵ_t. The following lemma is established to estimate the training noise rate η_t and hypothesis classification error ϵ_t in each iteration.

Lemma 3.2. *The training noise rate η_t and hypothesis classification error ϵ_t can be estimated assuming we have access to any two of the following quantities:*

a) *(Prior information) the number of negative samples, namely $|L_{-1,\checkmark}^t| + |L_{+1,\times}^t| + |U_{-1}^t|$, or $|L_{-1,\checkmark}^t| + |L_{+1,\times}^t|$ for the labeled set*
b) *(Type I or II error) the misclassification rate for either the positive or negative samples, namely $\frac{|L_{-1,\times}^t|}{|L_{-1,\times}^t| + |L_{+1,\checkmark}^t|}$ or $\frac{|L_{+1,\times}^t|}{|L_{+1,\times}^t| + |L_{-1,\checkmark}^t|}$.*

The set of equations for estimation are derived in the appendix.

Theorem 3.1 provides the relationship among the number of training samples, n, the training noise bound, η, and the classification error rate, ϵ, of the hypothesis that minimizes the training error. Lemma 3.2 offers a method to calculate the classification noise rate η_t in the t round. Inspired by Zhou and Li [33] and Goldman and Zhou [34], we state the following theorem that guarantees that classification performance will improve in each round.

Theorem 3.3. *The gap between the learned and optimal hypotheses as shown in PAC property (5) will decrease with high probability in each iteration with suitable sampling rates, α_{-1} and α_{+1}, whenever the following condition is satisfied:*

$$\left(|L_{-1}^{t+1}| + |L_{+1}^{t+1}|\right)\left(1-2\eta_{t+1}\right)^2 > \left(|L_{-1}^{t}| + |L_{+1}^{t}|\right)\left(1-2\eta_{t}\right)^2 \tag{7}$$

where $\left(|L_{-1}^{t+1}| + |L_{+1}^{t+1}|\right)$ is the total number of samples in the training set in round $t+1$, and η_{t+1} is the training noise rate.

Theorem 3.3 suggests the **stopping indicator** as follows:

$$\mathbf{1}\{\left(|L_{-1}^{t+1}|+|L_{+1}^{t+1}|\right)\left(1-2\eta_{t+1}\right)^2 \leq \left(|L_{-1}^{t}|+|L_{+1}^{t}|\right)\left(1-2\eta_{t}\right)^2\} \tag{8}$$

which evaluates to 1 when the condition in (7) is not satisfied. Frequent occurrences of such events might be a strong indication to stop the algorithm and avoid potential deterioration, as illustrated in § 4. The proof for Lemma 3.2 and Theorem 3.3 can be found in the appendix.

3.2.2 Surrogate Loss

The MIT proposed in the previous section initializes the training data based on common schedules, and iteratively refine the labels until a stopping condition is met. As suggested by Natarajan et al. [31], another approach is to recognize the presence of noise with $\rho_{-1} = \mathbb{P}(\tilde{y} = +1|y = -1)$ and $\rho_{+1} = \mathbb{P}(\tilde{y} = -1|y = +1)$ as the label corruption probability, and a design loss function as a substitute for the original loss to "reverse" the corruption process. The procedure of the surrogate loss approach is described below, and Algorithm 1 illustrates the implementation.

1) *Initialization:* Set the collected data where the presence is estimated by prior knowledge, i.e., occupancy schedules (line 1).
2) *Cross-validation:* For each parameter $\theta \in \Theta^{CV} \subset \Theta$ in a finite subset of the parameter space, obtain the empirical risks associated with θ, $\hat{R}_{CV}(\theta)$, following the standard cross-validation procedure on the surrogate loss, $\tilde{l}(t, y; \theta)$ (lines 2, 3).
3) *Learning with surrogate loss:* Identify the best parameter as suggested by the previous step, $\theta^* = \arg\min_{\theta\in\Theta^{CV}} \hat{R}_{CV}(\theta)$, and use it to obtain the new presence labels for the dataset (lines 4 to 6).

The first step in the above procedure is based on the multiview assumption, and the second step is motivated by parameter optimization $\theta = \{\rho_{+1}, \rho_{-1}\}$, since the conditional noise rates, $\rho_{\pm 1}$, are unknown.

Based on corrupted learning [35] and label-dependent loss [31], two different surrogate losses are summarized below:

- With the learning under corruption scheme, the surrogate loss is given by:

$$\tilde{l}(f(\boldsymbol{x}), \tilde{y}) = \frac{(1-\rho_{-\tilde{y}})l(f(\boldsymbol{x}), \tilde{y}) - \rho_{\tilde{y}} l(f(\boldsymbol{x}), -\tilde{y})}{1-\rho_{+1}-\rho_{-1}} \tag{9}$$

  where $\rho_{\tilde{y}}$ is the conditional noise rate, and $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is the original loss function.

Algorithm 1: Learning with Surrogate Loss

`Learning_Surrogate_Loss(`X, Prior, MaxIter`)`
Input: X: feature matrix of size $n \times m$, where n is the number of samples, m is the number of views.
Prior: expert knowledge for initialization.
Initialization:
1 $L \leftarrow$ Prior(X) `// initial estimation by` Prior
Cross-validation:
2 **for** $\theta \in \Theta^{CV}$ **do**
3 $\quad R_{CV}(\theta) \leftarrow$ CVEstMdl$(X, L, \tilde{l}(t, y; \theta))$
Learning with surrogate loss:
4 $\theta^* = \arg\min_{\theta\in\Theta^{CV}} R_{CV}$
5 EstMdl $\leftarrow$ `MdlEst`$(X, L, \tilde{l}(t, y; \theta^*))$
6 $L^* \leftarrow$ MdlPred$(X,$ EstMdl$, \tilde{l}(t, y; \theta^*))$
Output: L^* `// labeled datasets`

- The γ-weighted label-dependent loss is given by:

$$l_\gamma(f(\boldsymbol{x}), \tilde{y}) = (1-\gamma)\mathbf{1}(\tilde{y}=1)l(f(\boldsymbol{x}), \tilde{y}) + \gamma\mathbf{1}(\tilde{y}=-1)l(-f(\boldsymbol{x}), \tilde{y}) \tag{10}$$

  where γ is the weight chosen according to the conditional noise rates $\frac{1-\rho_{+1}+\rho_{-1}}{2}$.

The two surrogate loss functions above are designed such that the procedure of seeking the empirical risk minimizer (1) under the "corrupted" distribution is as if we are working with the original loss function under the "clean" distribution, where the labels are the ground truth (the derivation of (9) is obtained in the Appendix). The next theorem provides an upper bound on the regret (2) using the surrogate loss function (9).

Theorem 3.4. *[31] Let $l(t, y)$ be L-Lipschitz in t for every y, then with probability at least $1-\delta$, and $\hat{f} = \arg\min_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \tilde{l}(f(\boldsymbol{x}_i), y_i)$ be the ERM with the corrupted data,*

$$\delta R_{l,\mathcal{F}}(\hat{f}) \leq 4L_\rho \mathcal{R}(\mathcal{F}) + 2\sqrt{\frac{\log(1/\delta)}{2n}} \tag{11}$$

where $\mathcal{R}(\mathcal{F}) = \mathbb{E}_{\boldsymbol{x}_i,\epsilon_i}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(\boldsymbol{x}_i)\right]$ is the Rademacher complexity of the function class $\mathcal{F}$ with ϵ_i as the i.i.d. Rademacher (symmetric Bernoulli) random variables [36], and $L_\rho \leq 2L/(1-\rho_{+1}-\rho_{-1})$ is the Lipschitz constant of $\tilde{l}$ given in (9).

Theorem 3.4 suggests that the upper bound on regret decreases as we have more samples, whose infimum depends on the Lipschitz constant and complexity of the function class only. As the proposed procedure is like "exploration in the darkness", the result offers performance guarantee; nevertheless, the precondition is specified that neither $\rho_{\pm 1}$ is greater than .5. Indeed, the bound improves as the conditional noise rates reduce. We will refer to (9) as the **unbiased loss** (U. L.), since the expectation of the original loss $l(\cdot, \cdot)$ under the "clean" distribution is identical to that of the surrogate loss $\tilde{l}(\cdot, \cdot)$ under the "corrupted" distribution [35].

The **γ-weighted label-dependent loss** is designed in a similar fashion, except that the corrupted risk, $R_{l_\gamma,\mathcal{F}}(f)$, now is an affine transformation of the original risk, $R_{l,\mathcal{F}}(f)$:

Lemma 3.5. *[31] There exists a constant B that is independent of f such that by choosing $\gamma = \frac{1-\rho_{+1}+\rho_{-1}}{2}$ and $A_\rho = \frac{1-\rho_{+1}-\rho_{-1}}{2}$, and for all functions $f \in \mathcal{F}$,*

$$R_{l_\gamma,\mathcal{F}}(f) = A_\rho R_{l,\mathcal{F}}(f) + B \tag{12}$$

Intuitively, the loss puts more weights on data with labels that are less corrupted. With the chosen γ, optimization with $\tilde{l}_\gamma$ is equivalent to that with the original loss due to the affine relation (12). The next theorem gives performance guarantee.

Theorem 3.6. *[31] Let L be the Lipschitz constant for l as before, and $\hat{f} = \arg\min_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \tilde{l}_\gamma(f(\boldsymbol{x}_i), y_i)$, then with probability at least $1-\delta$,*

$$\delta R_{\tilde{l}_\gamma,\tilde{D},\mathcal{F}}(\hat{f}) \leq 4L\mathcal{R}(\mathcal{F}) + 2\sqrt{\frac{\log(1/\delta)}{2n}} \tag{13}$$

where $\mathcal{R}(\mathcal{F})$ is the Rademacher complexity.

There are two keys for using the surrogate loss approach to solve the NL problem. One, the initialization of the training set is based on a common occupancy schedule, which provides useful information for human behavior mining. Two, the surrogate loss is designed such that the expectation of the objective under the corrupted distribution is equivalent to that of the original problem. Hence, the solution is unaffected by the noise introduced in the initial phase.

3.3 Transfer Learning Methods

The main issue with the TL problem as outlined in § 2.2 is data abundance/scarcity. While the dataset from the households of interests, or target, is limited, the dataset from the different households that might have different energy usage patterns, or source, is ample.

The key lies in the answer to the question, how we can we ''apply the knowledge" gained from analyzing a dataset with similar characteristics to the dataset in question? We can completely ignore the question by performing BL, such as SVM, for each individual independently; however, it might be difficult for the learner to generalize the learning instances by relying on the limited data from the target household alone. In this case, data from the other households that share similar consumption patterns might be useful in estimating the correspondence between the occupancy and power measurements, crucial to the detector's performance overall.

For SVM, the learner assumes the form $f^{(k)}(\boldsymbol{x}) = \text{sgn}\left(\langle \boldsymbol{x}, \boldsymbol{\omega}_k\rangle\right)$, where $\boldsymbol{x} \in \mathcal{X}$ is the feature input and $\boldsymbol{\omega}_k$ is the parameter for individual k. Instead of training $\boldsymbol{\omega}_k$ independently, Evgeniou and Pontil [37] proposes the following decomposition:

$$\boldsymbol{\omega}_k = \boldsymbol{\omega}_0 + \boldsymbol{v}_k \tag{14}$$

where $\boldsymbol{\omega}_0$ is the common component that is shared in the group, and $\boldsymbol{v}_k$ represents the individual effect which is unique for each household. With the model specified above and based on the original problem as in (1), the new optimization is formulated as:

$$\begin{aligned} \min_{\boldsymbol{\omega}_0,\boldsymbol{v}_k,\xi_i^{(k)}} \quad & J_1(\xi_i^{(k)}) + J_2(\boldsymbol{\omega}_0, \boldsymbol{v}_k) \\ \text{s. t.} \quad & y_i^{(k)}\left\langle \boldsymbol{\omega}_0 + \boldsymbol{v}_k, \boldsymbol{x}_i^{(k)}\right\rangle \geq 1 - \xi_i^{(k)} \\ & \xi_i^{(k)} \geq 0, \;\; k \in S_T \cup \{K\}, i \in [m(k)] \end{aligned} \tag{P1}$$

where $m(k)$ is the data size for task k, $[m(k)] = \{1, ..., m(k)\}$ uses our set notation, $S_T \cup \{K\}$ is the union of the source (S_T) and target (K) index set, $J_1(\xi_i^{(k)})$ is the classification cost:

$$J_1(\xi_i^{(k)}) = \sum_{i=1}^{m(K)} \xi_i^{(K)} + \lambda_0 \sum_{k\in S_T}\sum_{i=1}^{m(k)} \xi_i^{(k)} \tag{15}$$

with λ_0 as the importance weight of the source tasks, and

$$J_2(\boldsymbol{\omega}_0, \boldsymbol{v}_k) = \frac{\lambda_1}{K} \sum_{t\in S_T\cup\{K\}} \|\boldsymbol{v}_k\|^2 + \lambda_2\|\boldsymbol{\omega}_0\|^2 \tag{16}$$

is the regularization term to avoid overfitting. There are two trade-offs in this formula:

- Source vs. target tasks
- Individuality vs. conformity

The individuality vs. conformity trade-off effectively controls the amount of knowledge transfer across tasks [37]. In fact, setting $\lambda_1/\lambda_2 \to \infty$ would correspond to learning tasks independently, or no knowledge transfer, and setting $\lambda_1/\lambda_2 \to 0$ would result in learning a single model, or transfer all knowledge.

New in this study, the source vs. target trade-off recognizes the relative importance of learning the target task well, as compared to the source, by setting $\lambda_0 < 1$. Since the target dataset is usually smaller than the source dataset, this avoids the target from being overwhelmed, and can further improve the performance. The principles of this tradeoff is illustrated in § 4.

The common and individual models of (P1), $\boldsymbol{\omega}_0^*$ and $\boldsymbol{v}_k^*$, are related as follows:

Lemma 3.7. *It can be shown that $\boldsymbol{\omega}_0^* = \frac{\lambda_1}{K\lambda_2}\sum_{k=1}^{K} \boldsymbol{v}_k^*$.*

The connection of (P1) to the standard SVM is depicted in the next proposition (see the appendix for the proof):

Proposition 3.8. *With the following transformation of features, Φ, for samples in all tasks, $k \in \{1, ..., K\}$,*

$$\Phi(\boldsymbol{x}, k) = \left(\frac{\boldsymbol{x}}{\sqrt{\mu}}, \underbrace{\boldsymbol{0}, ..., \boldsymbol{0}}_{k-1}, \boldsymbol{x}, \underbrace{\boldsymbol{0}, ..., \boldsymbol{0}}_{K-k}\right) \tag{17}$$

and the construction of the parameters,

$$\boldsymbol{\omega} = (\sqrt{\mu}\boldsymbol{\omega}_0, \boldsymbol{v}_1, ..., \boldsymbol{v}_K) \tag{18}$$

where $\mu = \frac{K\lambda_2}{\lambda_1}$, (P1) can be converted to the standard SVM.

According to Proposition 3.8, we can use the popular SVM solvers, e.g., LibSVM [28], to tackle the TL problem. Also, we can extend the proposition to nonlinear SVM by "lifting" $\Phi(\boldsymbol{x}, t)$ to higher, even infinite, dimensions, which

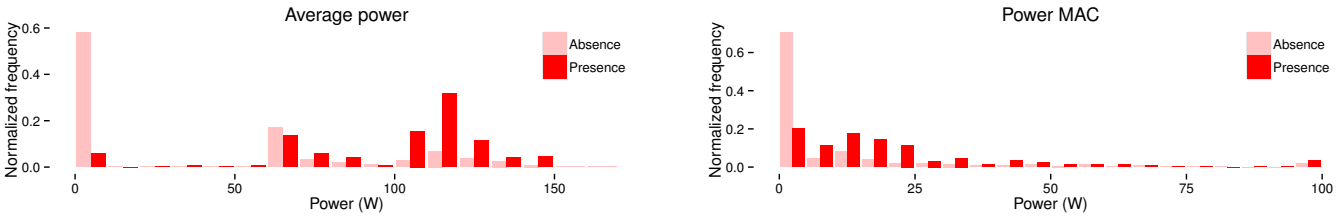


Fig. 3: Conditional distributions of features, i.e., average power and MAC, given occupancy states for u17.

corresponds to various kernels, e.g., radial basis and sigmoid, in the dual formulations. In the experimental section, we will illustrate the empirical rules for selecting μ, and our strategy to implement λ_0 for the source-target trade-off.

3.4 Power Features Exploration

The key question resolved in this section is: which features based on power usage are the most indicative of occupancy? We motivate the exploration by observing the power traces in tandem with the occupancy information for residential and commercial buildings (Fig. 2).

Although the presence of occupants generally means a higher power usage, it is not uncommon for occupants to leave their computers on while they are out of office. In the next section, power features are designed based on power magnitude, transition, and transient effects. We also investigate how the length of aggregation window and the electricity sampling rates influence the separability of these features. Throughout this section, x_i is used to denote real power measurement at time index i.

3.4.1 Power Magnitude

When people are in their homes or offices, they tend to use more devices, such as computers, televisions, microwaves, and lights than when they are absent. This indicates that the distributions of power magnitudes, as shown in Fig. 3, are distinct for the states of occupancy and vacancy.

Specifically, the mean of the distribution for occupancy is higher than the mean for the vacancy, and it also has more spreads over the spectrum due to the variety of devices used. Additionally, there are more high power samples due to devices like microwaves, kettles, blenders, etc., that consume a large amount of power in a short period of time.

The overlap between the two distributions can be attributed to two factors. One, there are devices such as televisions that are in stand-by modes, or devices such as refrigerators that are always on regardless of human presence. Two, individuals forget to turn off devices such as computers, lights, and HVAC when they leave the rooms, resulting in energy waste. Clearly, features that look beyond power magnitudes are required to distinguish occupancy presence and absence.

3.4.2 Transition Effect

The basic *learning problem* is described by $(l, \mathbb{D}, \mathcal{F}, e_n)$, where $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is the *loss function* to penalize misdetection, e.g., the 0-1 loss $l_{01}(y, f(\boldsymbol{x})) = \mathbf{1}(y \neq f(\boldsymbol{x}))$, which evaluates to 1 if the two arguments differ and 0 otherwise.

A reliable indication of human presence is when individuals switch devices ON or OFF, causing an abrupt change in power consumption. This transition effect has been used for non-intrusive load monitoring (NILM) to disaggregate devices from power measurement [38]. Also, change-point detection methods in time series also analyze the divergence of distributions to signal transitions in states.

The transition effect is characterized by the following features:

- Maximum absolute change (MAC):

$$\delta_{MAC} = \max_{i,j \in S_d} |x_i - x_j| \tag{19}$$

  where S_d is the set of points with window size d.
- Counts of ON/OFF events (SOF), δ_{SOF}, is the number of events such that: the change of power $|x_i - x_{i-1}|$ is greater than δ_p, and no such changes occur in the next δ_t period.

Distribution of MAC is shown in Fig. 3, which illustrates separability in the new dimension.

3.4.3 Transient Effect

Apart from transitions that happen infrequently, a critical observation from Fig. 2 is that despite the high power level, there are much fewer "spikes" and "ripples" during human absence. One plausible explanation is that when devices, such as computers, are in active use, they frequently switch among several modes, i.e., browsing, simulation, and word processing. Compared to the stand-by mode with a constant power need, these various modes have different power needs.

Features are designed to capture the transient effect:

- Mean of absolute difference (MAD):

$$\delta_{MAD} = \frac{1}{d} \sum_{i \in S_d} |x_i - x_{i-1}| \tag{20}$$

- Mean of absolute height difference (MAHD):

$$\delta_{MAHD} = \frac{1}{|C_d|} \sum_{(i,j) \in C_d} |x_i - x_j| \tag{21}$$

- Standard deviation (SD):

$$\delta_{SD} = \sqrt{\frac{1}{d-1} \sum_{i \in S_d} (x_i - \bar{x})^2} \tag{22}$$

where $C_d = \{(i,j) | x \text{ monotone between } i, j, i < j \in S_d\}$, and $\bar{x} = \frac{1}{d} \sum_{i \in S_d} x_i$ is the average power.

As shown in Fig. 3, the distributions of δ_{MAD} and δ_{SD} achieve good feature separability. Although it illustrates the active/inactive device usage, this information indirectly relates to occupant presence, requiring corroboration with other features for the occupancy inference.

3.4.4 Window Effect

The "window" mentioned above is used to aggregate a chunk of data to calculate features. The choice of window size, d, depends on two independent factors: the time resolution of occupancy detection, and the separability of obtained features. Although many authors use the former to determine window size [15], [16], [17], [24], the latter is critical to detection performance, and should be studied systematically.

The more separable the feature, the easier it is to classify [27]. Several divergence metrics for distributions have been employed to measure feature separability, including the Jensen-Shannon (JS) divergence, Hellinger distance (HD), total variation (TV) distance, and Bhattacharyya distance (BD). These definitions are provided in the appendix. For these metrics, larger values indicate that the features are more effective in distinguishing between occupancy and vacancy states, key to the detection accuracy.

While average power is generally insensitive to window size, other features that characterize the transition and transient effects are more sensitive to window size and have better qualities as the window expands, assuming the same sampling rates (Fig. 4). Since occupancy resolution decreases for large windows, a sampling rate of 1 sample per minute during the 15 minutes period (i.e., 15 aggregated data points) seems to be a good trade-off.

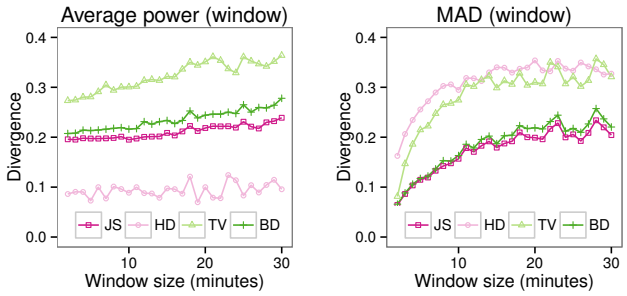


Fig. 4: Divergence measures to study the window effect. HD and TV are offset by -1 and -0.25 for comparison.

3.4.5 Sampling Effect

The sampling rate refers to how frequently power is monitored. While the window size determines the span of information to investigate, the sampling rate controls the density of data to collect. Higher sampling rate means denser sample time data.

Data collected at higher rates usually reveal more information, e.g., the use of high frequency monitoring for power disaggregation [39]. But, it also introduces redundancy, especially for the detection of occupancy that typically changes slowly. There are limitations to data storage and transmission bandwidth, as well as trade-offs between privacy and detection resolution at different sampling rates.

Using the divergence metrics, we investigate the separability of the features as we increase the sampling intervals, with a fixed window size of 30 minutes (Fig. 5). Although the power average feature seems to be relatively robust when sampling rates decrease, the MAD feature becomes less effective in distinguishing between presence and absence states. This phenomenon can be observed from the diminishing divergence metrics. In general, features that describe the transition and transient effects are most impacted by a sparse power usage request, leaving out valuable temporal information as a result. The sampling frequency is 1 sample per second for the ECO [15] and UMass Smart* [14] datasets, and 1 sample per minute for the PC [23] dataset used for evaluation. In practice, it is sufficient to capture the transition and transient effects with **1 minute sampling interval**.

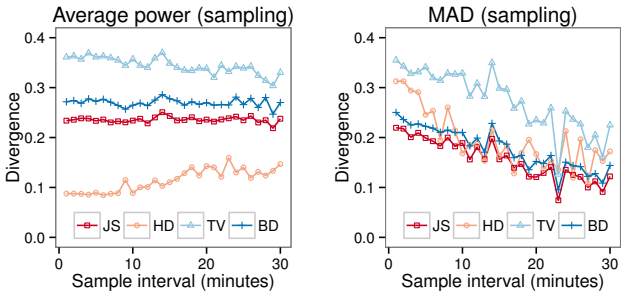


Fig. 5: Divergence measures for power average and MAD to study the sampling effect.

4 Experimental Evaluation

Using the datasets for office cubicles (PC) [23],and the ECO [15] and Smart* [14] that cover the commercial and residential buildings, we evaluate the BL, NL, and TL methods and facilitate a comparison with previous work [14], [15].

4.1 Experimental Setup and Data Collection

4.1.1 Personal Cubicles in an Office

The goal is to evaluate individual presence detection based on power measurement in personal cubicles to facilitate lighting and HVAC control for social games [19]. Located in Cory Hall on UC Berkeley campus, the office is instrumented with the following sensors [23]:

Ultrasonic sensors, placed on user's desk as shown in Fig. 6, measures the distance to their nearest obstacles, i.e., individuals or chairs, by recording the time of flight Δt of triggered sound wave: $d_{obs} = \frac{1}{2}\Delta t v_s$, where $v_s \doteq 340\text{m/s}$ is the velocity of sound in the air. Radio modules, controlled by Arduino microprocessors, implement the ZigBee networking protocol for data transmission (Fig. 7). Measurements can be processed by simple thresholding for presence inference.

Acceleration sensors, attached to the chairs (Fig. 6), detects presence by sensing chair motions. Collected by our previously developed Building-in-Briefcase (BiB) sensors [40], the data is stored on our PI server for remote access.

WiFi access points, deployed across the office for indoor positioning by measuring the radio signal strength (RSS) [41], can be used to infer the presence of smart phones when they are connected to the network. Due to the IEEE 802.11 protocols, the connection interval ranges from several seconds to increments of ten minutes for energy efficiency.

In addition, we asked 4 users (id: 8, 17, 20, 26) to complete surveys that indicate their presence designed with 5 minutes resolution. This user survey data is combined with the above sensor measurements to provide ground truth. The power consumption is collected by ACme sensors [42]

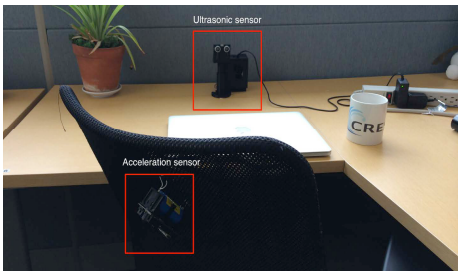


Fig. 6: Experimental setup of presence sensing networks.

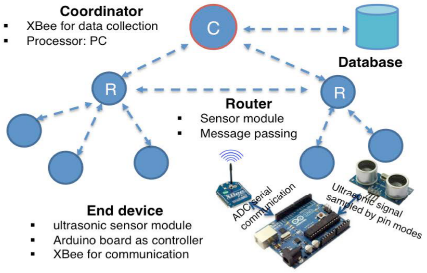


Fig. 7: Communication configuration based on ZigBee.

at a resolution of a second. More details can be found in our previous work [23]. We refer to this dataset as "PC" in the discussions.

4.1.2 Residential Buildings

For the residential building occupancy detection, we use the publicly available Electricity Consumption and Occupancy (ECO) dataset, which consists of fine-grained electricity and occupancy measurements for five Swiss households (id: r1, ..., r5) during summer (July to September 2012) [15]. We also employ the UMass Smart* home dataset [14], [43] from two homes during the summer in western Massachusetts.

Aggregate power is collected by off-the-shelf digital electricity meters at a sampling rate of 1 Hz for ECO and Smart*. Occupancy information is entered manually by residents using the tablet mounted near the main entrance. Details about the households (number of occupants, types of devices, etc.) and data preprocessing techniques are described in [14], [15], [43].

4.2 Base Learning Results

We analyze the BL problem as described in § 2.2 by examining both residential buildings (ECO) and personal cubicle (PC) data. Features, including average power, MAC, SOF, MAD, MAHD, SD, are derived within a 15 minutes window, and the sampling intervals are 1 second for ECO and Smart* and 1 minute for PC.

Hypothesis 1. *Power can be used to infer occupancy.*

First, we verify this hypothesis using simple thresholding methods. Then, we demonstrate in Figs. 8 and 9 for PC and ECO that the accuracy improves with more advanced classifiers such as J48 decision tree and random forest.

Power thresholds are based on magnitudes (Mag/Th), changes in power magnitude (Chg/Th), and changes in percentage (Prc/Th), where the thresholds can be optimized over the training set, as detailed in [23]. We also used the static schedule as the baseline, which indicates occupancy or

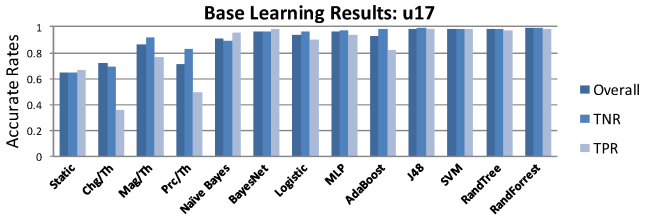


Fig. 8: BL results for u17 in the PC dataset, obtained by 10-fold cross-validation. The methods, except for the static baseline, Chg/Th, Mag/Th, Prc/Th, are implemented by the Weka Machine Learning Toolkit [29].

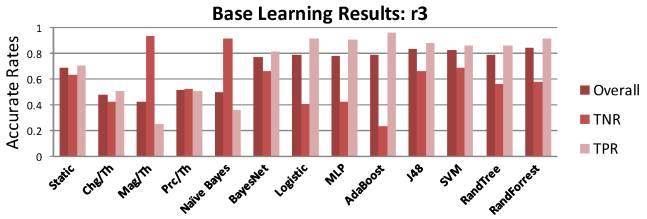


Fig. 9: BL results for household r3 in the ECO dataset by 10-fold cross-validation.

vacancy from 8am to 6pm for the PC or ECO, and vacancy or occupancy for the rest of the day.

Our findings show that power thresholding can achieve basic inference. As for accuracy of power features, Mag/Th is more suitable for PC and Prc/Th is more suitable for ECO. As PC has a few devices such as computers and desk lamps that are turned on or off when individuals arrive or leave their offices, Mag/Th are sound indicators of occupancy. As ECO has loads that are in stand-by modes even during individuals' absence, Mag/Th is not a suitable indicator. Since some devices such as microwaves, kettles, that draw high instantaneous power during individuals' presence, Prc/Th are strong indicators for occupancy. Since the households follow regular occupancy patterns, the static baseline produces effective results, and even outperforms the power thresholding methods in ECO.

The advanced classifiers, especially SVM, random forest, and J48 decision tree, seem to outperform the baseline methods for overall accuracy, true negative rates (TNR) for accurate vacancy inference, and true positive rates (TPR) for correct occupancy inference. Due to devices in stand-by modes and devices with auto-triggers [15], the advanced classifiers' performance is slightly worse for ECO. However, the experimental results clearly show that power is a reliable indicator for occupancy.

Hypothesis 2. *Inference accuracy is limited by training sample size, and improves as more data is available.*

To evaluate this hypothesis, we show the relationship between the inference accuracy for well-performing classifiers and training data size in Fig. 10. The graph illustrates that the classifiers' accuracy improves rapidly and reaches the upper limit with 5% or more training data. However, as the training data size falls below 5%, or 1 to 4 days of occupancy data, the classifiers' accuracy deteriorates significantly. The key problem is that while the learner can have a reasonable estimate of the power-occupancy correspondence with 5% of data, it seems difficult for the learner to generalize the learning instances with 1% or less data. But, if we can train

a well-performing classifier with only 1% or less data, more than 80% of the survey effort can be saved for the user. In the next section, we will explore this extreme data sparsity regime, or NL and TL problems.

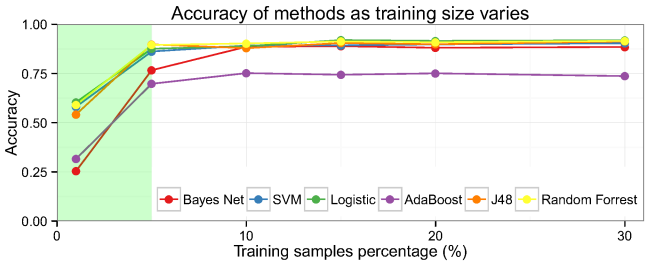


Fig. 10: Accuracy of well-performing classifiers implemented for r2 in ECO, as training set size varies in the range of (1, 5, 10, 15, 20, 30)%.

4.3 Non-Intrusive Learning Results

While the previous section discusses occupant presence detection using readily available energy data, this section examines the following hypothesis:

Hypothesis 3. *When no training data is available, our power-based occupancy detectors can still work by exploiting the common schedules of average users.*

Evaluation of this hypothesis requires the examination of the two paradigms proposed in § 3.2:

- Multiview-based Iteration (MIT): initialized by time schedules, the presence labels are iteratively refined until certain stopping conditions are met.
- Recognize the noise in the training data, and learn under corruption by designing surrogate losses.

Results for PC and ECO are reported in Figs. 11 and 12 for MIT, unbiased loss (U. L., (9)), and γ-weighted loss (10).[2] Generally, all methods deliver satisfactory performances compared to the static schedule. However, due to the variability in the occupancy and device usage, ECO is more challenging than PC. According to TNR and TPR, the γ-weighted loss seems to be a better absence detector, and unbiased loss excels at presence detection.

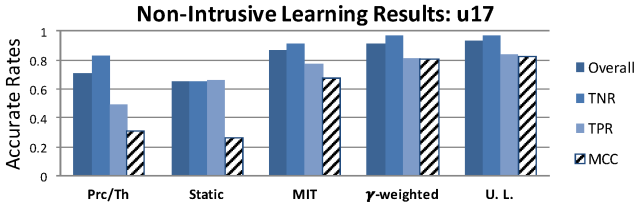


Fig. 11: NL results for user u17 in the PC dataset by 10-fold cross-validation for baseline methods and MIT, γ-weighted, and unbiased losses.

Similar to BL results (Fig. 9), the TNR/TPR rates report for ECO (Fig. 12) indicates that the NL methods have

2. As illustrated in Fig. 13, we implemented Naïve Bayes in each iteration of MIT, and terminated the iteration based on the stopping conditions. LibSVM is used to realize the γ-weighted loss with radial basis function (RBF) kernel. The logistic loss is selected for the unbiased loss, optimized by Nesterov's accelerated gradient method [44].

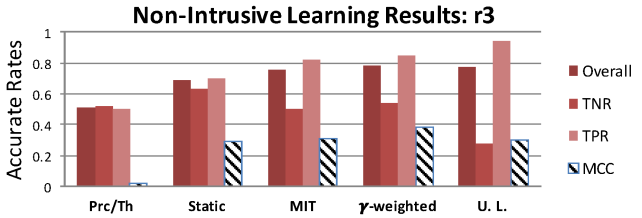


Fig. 12: NL results for household r3 in the ECO dataset by

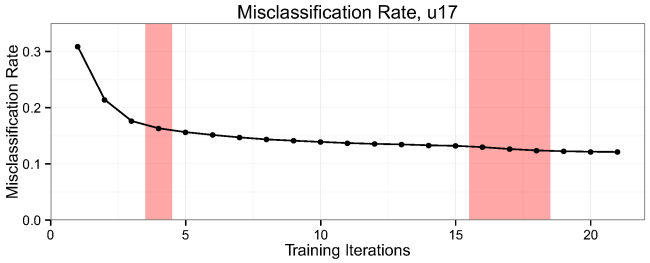


Fig. 13: Misclassification rate during training iterations of MIT for u17 in PC. As the user can only observe the stopping conditions, or the red regions, the user can terminate the training to avoid deterioration [23].

relatively low TNR, or more mistakes when the households are vacant. This is mainly due to: (1) user behavior – sometimes the power consumption is high or has considerable fluctuations despite user absence, (2) the balance of dataset – users in some ECO households tend to be present in late mornings or early afternoons, as shown in the example traces (Figs. 2 and 14), resulting in more instances of occupancy than vacancy data [15]. While (1) illustrates the fundamental limitation in the proposed approach, or the difficulty in using only power data to indicate user presence or absence due to similar consumption patterns, (2) can be resolved by using existing methods such as putting a larger penalty on vacancy misclassification [45] to improve the TNR performance.

We employ the Matthews correlation coefficient (MCC), as suggested in [15] as a balanced measure of the prediction quality to overcome the difficulties in comparing different sizes of positive and negative instances:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

where TP, TN, FP, FN are the numbers of true positive, true negative, false positive (ground truth: vacancy, estimation: occupancy), and false negative (ground truth: occupancy, estimation: vacancy) instances. The MCC returns a value between -1 and +1: a coefficient of +1 represents a perfect prediction, a coefficient of 0 represents no better than random prediction, and a coefficient of -1 indicates total disagreement between prediction and observation. Overall, the proposed NL methods significantly outperform the baseline models in the PC dataset, but only slightly outdo the baseline models in the ECO dataset. The latter is likely caused by a lack of substantial discrepancy of power characteristics between occupancy and vacancy states in the residential buildings of ECO. Based on the PC results which measure the plugloads in an office cubicle, further improvements can be achieved by sensor fusion [10] or submetering key devices such as plugloads or lighting systems.

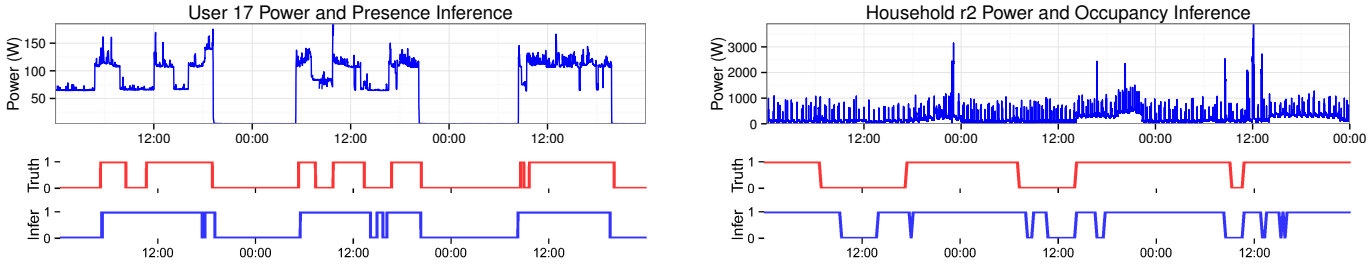


Fig. 14: Examples of presence detection for u17 (left) and r2 (right) with MIT and γ-weighted loss, respectively. The power traces are shown on the top, whereas the bottom plots the true (red) and estimated (blue) occupancy, for comparisons.

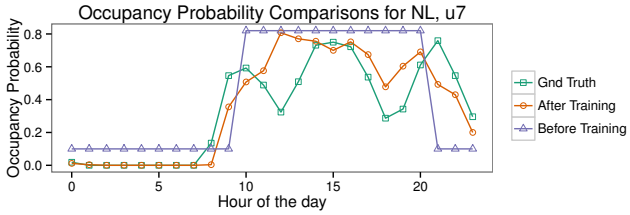


Fig. 15: Occupancy schedules for NL include the shared profile (green), the learned one by MIT (blue), and the ground truth (red) for u17 in PC.

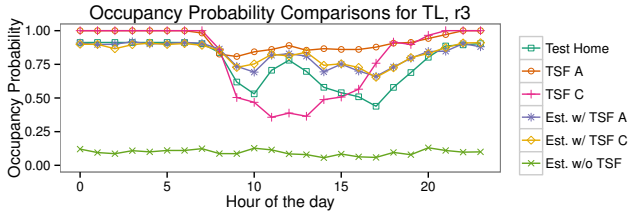


Fig. 16: Occupancy schedules for TL include the estimation w/ (purple) and w/o (green) transfer, and the ground truth for the target (red) and source (blue) households.

Examples of occupancy detection in time series are demonstrated in Fig. 14. This data suggests that the detectors base their occupancy determinations on power magnitudes as well as transition and transient effects. Compared to thresholding methods, the γ-weighted loss is more effective in occupancy detection, especially for the ECO dataset with periodic power surges and diversified device usage patterns. In Fig. 15, hypothesis 3 is further verified. Closely following the ground truth, the learned occupancy schedule outperforms the common template for occupancy detection.

4.4 Transfer Learning Results

As depicted in Scenario 2 in § 1, the TL problem is due to limited data from new individuals, or target, and abundant data from similar individuals, or source. The following hypothesis resolves this data imbalance:

Hypothesis 4. *Occupancy detection for the target can be improved by incorporating data from the source.*

Proposed in § 3.3, the model consists of common and individual components, and the solution is developed by solving the optimization program (P1). The source-target tradeoff parameter is set to balance the number of samples in the two sets (e.g., if there are n_S and n_T samples in the source and target, respectively, then $\lambda_0 = n_S/n_T$). While the conformity-individuality parameter μ is usually set at 0 to 2, a lower value encourages more transfer for conformity.

In Fig. 17, results of TL applied to ECO is reported as ECO is more challenging and relevant due to the prevalence of smart meters in homes. The schemes of Transfer A and B refer to the inclusion of one home as the source dataset (e.g., r1 as the source for r2 to r5, and r3 as the source for r1), or all other homes except for the target as the source dataset (e.g., r1 to r4 for target r5). Standard SVM performs poorly with only 1% of training data, which amounts to 30 to 70 randomly sampled points, with each sample representing 15 minutes occupancy/vacancy state. It is promising to note that due to abundant training data, both transfer schemes are highly accurate.

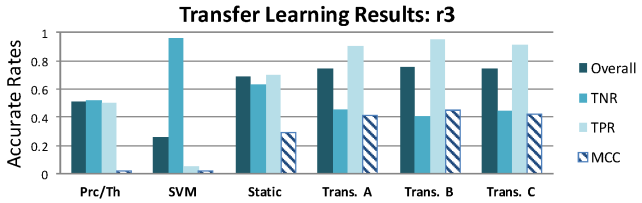


Fig. 17: TL results when only 1% of target dataset is available, averaged over 100 independent trials.

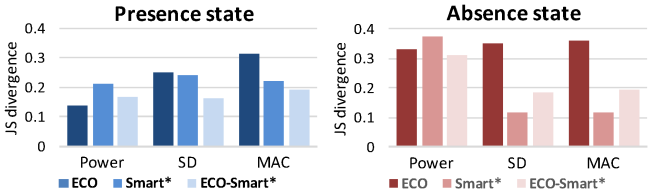


Fig. 18: JS divergence for normalized features of homes within (ECO, Smart*) and across (ECO-Smart*) datasets.

While the hypothesis presumes knowledge transfer from "similar" individuals, Fig. 16 illustrates that this requirement is far from stringent. Even though the source's occupancy schedule is very different from the target's schedule, the results of learning with transfer from source data is more favorable than the results of no transfer.

Further, we demonstrate TL's capability by transferring knowledge across two publicly available datasets (Transfer scheme C). Specifically, from the Smart* dataset reported by Chen et al. [14], we use two households' power usage information during the summer (Home A and B) to predict occupancy for homes in the ECO dataset. While the house-

Dataset	BL				NL				TL	
	HMM [15]	SVM-PCA [16]	HMM-PCA [16]	Rand. Forest	NIOM [14]	MIT	γ-weighted	U. L.	Tran. B	Tran. C
r1	0.83	0.83	0.83	**0.88**		0.74	0.83	0.83	0.83	0.82
r2	0.82	**0.92**	0.90	**0.92**		0.74	0.75	0.77	0.78	0.80
r3	0.81	0.83	0.82	**0.84**		0.76	0.78	0.77	0.76	0.75
Home A				**0.92**	0.79	0.81	0.78	0.84	0.87	0.84
Home B				**0.96**	0.91	0.85	0.84	0.89	0.91	0.86

TABLE 1: Comparison of our results with prior art [14], [15], [16], showing the overall accuracy metric. While the best performance in BL is bolded, the best performances in NL and TL are underlined.

holds in Smart* and ECO differ in geographical locations and home sizes, they share similar energy consumption patterns. This is verified by the JS divergence metrics for power features conditioned on the presence/absence states within and across datasets (Fig. 18). Transfer scheme C's performance is on par with schemes A and B in terms of the overall accuracy and MCC (Fig. 17). The learned occupancy schedule for r3 in ECO is closer to the ground truth due to the similarity with the homes in Smart* (Fig. 16). The results indicate tangible benefits to preselecting sources for TL by segmenting customer energy consumption data [46].

4.5 Comparison with Prior Art

We compare the best results from previous work [14], [15], [16] that also use the ECO and Smart* datasets for evaluation. For BL, the methods include hidden Markov model (HMM) [15], HMM-PCA, and SVM-PCA [16], which use principal component analysis (PCA) for feature selections.

For NL, Chen et al. [14] proposed a threshold-based non-intrusive occupancy monitoring (NIOM) algorithm, which assumes constant presence during nighttime, and then clusters the occupancy based on the deviation from nighttime power features. One drawback with this approach is that the nighttime power usage used to set the thresholds may not be an accurate indicator for occupancy. We address this problem by two distinct ideas: MIT uses a rough occupancy schedule to initialize the data, and then refines the labels by exploiting the power information; the surrogate loss methods recognize the noisy labels in initialized data, and customize loss functions to reduce the adverse effects.

Due to a lack of previous work that also use the ECO and Smart* datasets, we reported our TL results without a comparison. The evaluation procedures for transfer B and C are reported in § 4.4. Since there are only 7 days of data for each household in Smart*, we randomly selected 10% of the target data, or less than one day of training effort, and combined it with data from households (r1 to r3) in ECO for training.

Table 1 indicates that the random forest has the highest overall accuracy among all competitors in BL. The unbiased surrogate loss (U.L.) outperforms NIOM for Home A (84% vs. 79% accuracy), but slightly underperforms for Home B (89% vs. 91%). However, it remains consistently competitive for ECO compared to MIT and γ-weighted surrogate loss method. TL within the same dataset (Tran. B) seems to be more efficient than TL across two datasets (Tran. C). Overall, both TL performances are satisfactory due to the transfer of shared patterns.

5 Related Work

Due to the large volume of relevant methodology and application literature, we focus on the following three topics that are most pertinent, and refer the readers to the references in § 3 and its bibliography.

Occupancy Detection. As for occupancy detection, on the one end of the spectrum are the more specialized but reliable devices such as PIR [7], magnetic reed switches [2], [3], cameras [3], telephone handset sensors [47], network logins and GPS trackers [47], active RFID badges [5], and smart phones [48]. On the other end of the spectrum are the mechanisms that safeguard user privacy but are more error-prone, such as models based on environmental measurements, including temperature [10], CO_2 [8], [9], particulate matter [11], and ultrasonic chirps [49]. Concerns for practical implementations often include system cost, faulty installations, maintenance, and required training period.

Energy Data Mining. The analysis of electricity data has been conducted to cluster households [50] to estimate socio-economic status [51], identify usage patterns for prediction [52], [53], and provide tailor-made information to the families [54]. LBNL used ACme system to measure a large building space and demonstrated reliable performance [55]. Non-intrusive load monitoring (NILM), pioneered by Hart [56], has also been popular. It infers device-level ON/OFF states from the aggregate signal based on various power features [38].

Power for Occupancy Detection. Molina et al. [13] demonstrated the capability of power for occupancy detection through visual inspection, while Dong et al. [10], Chen et al. [14], and Kleiminger et al. [15] quantitatively assessed the results of using power for occupancy detection. Kleiminger et al. [16] subsequently reported improvement from 82% to 90% for the ECO dataset [15] based on a comprehensive set of features from three power phases, which were processed by feature selection and principal component analysis (PCA). Yang et al. [17] also experimented with a range of classifiers, such as random forest, decision tree, multilayer perceptron, using the Weka Machine Learning Toolkit [29].

BL in § 3.1 is conducted in a similar fashion as BL in the antecedent literature as baselines. However, the major difference is our solution to the data scarcity problem, or NL, as an extension of our previous work [23], which carries out learning without collecting presence data, and TL, which exploits the potential for improvement by incorporating data from other sources.

6 Conclusion and Future Work

Occupancy detection for residential and commercial buildings alike is crucial to improving energy efficiency, user comfort, and space utility. The pervasiveness of electricity meters eliminates the additional system cost and setup/maintenance efforts. Hence, electricity meters are viable candidates for presence sensing, with the added benefits of safeguarding privacy information (compared to cameras), and improving reliability (compared to environmental measurements).

The potential of energy meters is thoroughly explored by studying data from both residential (ECO and Smart*) and commercial (PC) buildings. Various power features, e.g., magnitude, MAC, SOF, MAD, SD, are investigated. The capability of power for presence detection is first demonstrated in BL with methods like SVM and random forest. For non-intrusive learning without training data of occupancy, or NL, the proposed multiview-based iteration and surrogate loss approaches use common schedules as rough estimates to learn refined ones that are much closer to the ground truth. As a result, accuracy rates are 74 to 89% for residential buildings and about 90% for offices. TL approach tackles the case when data from other sources are used in the current learning task. Its results further confirm the appropriateness of using power to detect occupancy by producing superior performance as compared to standard SVM.

For the future work, it will be meaningful to extend NL and TL from indicating occupancy (classification) to estimating the number of people (regression). Additionally, it will be promising to perform sensor fusion with other mobile nodes, such as smart phones, fitness trackers, and automobiles, for further improvement. Last but not least, as shown in this paper's inferences, smart meters can be used to infer various user characteristics, raising the important issue of preserving user privacy while maintaining system functions, which are worth exploring.

Acknowledgments

This research is funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. The authors are also thankful to Dr. Wilhelm Kleiminger and Dong Chen for the access to the ECO and Smart* datasets and meaningful discussions, and Judy Huang and the reviewers for their effort to improve the manuscript.

References

[1] J. McQuade, "A system approach to high performance buildings," *United Technologies Corporation, Tech. Rep*, 2009.

[2] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse, "The smart thermostat: using occupancy sensors to save energy in homes," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2010, pp. 211–224.

[3] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng, "Occupancy-driven energy management for smart building automation," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. ACM, 2010, pp. 1–6.

[4] M. Jin, H. Zou, K. Weekly, R. Jia, A. M. Bayen, and C. J. Spanos, "Environmental sensing by wearable device for indoor activity and location estimation," *arXiv preprint arXiv:1406.5765*, 2014.

[5] J. Scott, A. Bernheim Brush, J. Krumm, B. Meyers, M. Hazas, S. Hodges, and N. Villar, "Preheat: controlling home heating using occupancy prediction," in *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 2011, pp. 281–290.

[6] R. F. Dickerson, E. I. Gorlin, and J. A. Stankovic, "Empath: a continuous remote emotional health monitoring system for depressive illness," in *Proceedings of the 2nd Conference on Wireless Health*. ACM, 2011, p. 5.

[7] T. A. Nguyen and M. Aiello, "Energy intelligent buildings based on user activity: A survey," *Energy and buildings*, vol. 56, pp. 244–257, 2013.

[8] K. Weekly, N. Bekiaris-Liberis, M. Jin, and A. Bayen, "Modeling and estimation of the humans' effect on the co2 dynamics inside a conference room," *Control Systems Technology, IEEE Transactions on*, vol. 23, no. 5, pp. 1770–1781, 2015.

[9] M. Jin, N. Bekiaris-Liberis, K. Weekly, C. Spanos, and A. M. Bayen, "Sensing by proxy: Occupancy detection based on indoor co2 concentration," in *The 9th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM'15), Nice, France*, 2015.

[10] B. Dong, B. Andrews, K. P. Lam, M. Höynck, R. Zhang, Y.-S. Chiou, and D. Benitez, "An information technology enabled sustainability test-bed (itest) for occupancy detection through an environmental sensing network," *Energy and Buildings*, vol. 42, no. 7, pp. 1038–1046, 2010.

[11] K. Weekly, D. Rim, L. Zhang, A. M. Bayen, W. W. Nazaroff, and C. J. Spanos, "Low-cost coarse airborne particulate matter sensing for indoor occupancy detection," in *Automation Science and Engineering (CASE), 2013 IEEE International Conference on*. IEEE, 2013, pp. 32–37.

[12] U. EIA, "How many smart meters are installed in the United States, and who has them?" http://www.eia.gov/tools/faqs/faq.cfm?id=108&t=3, [Online; accessed 12/2016].

[13] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building*. ACM, 2010, pp. 61–66.

[14] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, and P. Shenoy, "Non-intrusive occupancy monitoring using smart meters," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013, pp. 1–8.

[15] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, "Occupancy detection from electricity consumption data," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013, pp. 1–8.

[16] W. Kleiminger, C. Beckel, and S. Santini, "Household occupancy monitoring using electricity meters," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 975–986.

[17] L. Yang, K. Ting, and M. B. Srivastava, "Inferring occupancy from opportunistically available sensor data," in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*. IEEE, 2014, pp. 60–68.

[18] E. McKenna, I. Richardson, and M. Thomson, "Smart meter data: Balancing consumer privacy concerns with legitimate applications," *Energy Policy*, vol. 41, pp. 807–814, 2012.

[19] L. Ratliff, M. Jin, I. Konstantakopoulos, C. Spanos, and S. Sastry, "Social game for building energy efficiency: Incentive design," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, Sept 2014, pp. 1011–1018.

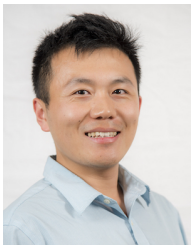
[20] C. Gomez and J. Paradells, "Wireless home automation networks: A survey of architectures and technologies," *IEEE Communications Magazine*, vol. 48, no. 6, pp. 92–101, 2010.

[21] M. Jin, N. Bekiaris-Liberis, K. Weekly, C. Spanos, and A. Bayen, "Occupancy detection via environmental sensing," *Transaction on Automation Science Engineering*, no. 99, pp. 1–13, 2016.

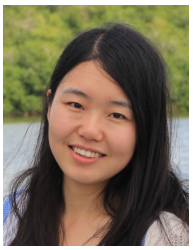
[22] M. Jin, W. Feng, P. Liu, C. Marnay, and C. Spanos, "Mod-dr: Microgrid optimal dispatch with demand response," *Applied Energy*, vol. 187, pp. 758–776, 2017.

[23] 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 *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, ser. BuildSys '14, 2014, pp. 1–10.

[24] G. Tang, K. Wu, J. Lei, and W. Xiao, "The meter tells you are at home! non-intrusive occupancy detection via load curve data," in *IEEE International Conference on Smart Grid Communications*, 2015.
[25] A. B. Brush, B. Lee, R. Mahajan, S. Agarwal, S. Saroiu, and C. Dixon, "Home automation in the wild: Challenges and opportunities," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. ACM, 2011, pp. 2115–2124.
[26] D. Kahneman, *Thinking, fast and slow*. Macmillan, 2011.
[27] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.
[28] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.
[29] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
[30] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the eleventh annual conference on Computational learning theory*. ACM, 1998, pp. 92–100.
[31] N. Natarajan, I. S. Dhillon, P. K. Ravikumar, and A. Tewari, "Learning with noisy labels," in *Advances in Neural Information Processing Systems*, 2013, pp. 1196–1204.
[32] L. G. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.
[33] Z.-H. Zhou and M. Li, "Tri-training: Exploiting unlabeled data using three classifiers," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 11, pp. 1529–1541, 2005.
[34] S. Goldman and Y. Zhou, "Enhancing supervised learning with unlabeled data," in *ICML*. Citeseer, 2000, pp. 327–334.
[35] B. van Rooyen and R. C. Williamson, "Learning in the presence of corruption," *arXiv preprint arXiv:1504.00091*, 2015.
[36] P. L. Bartlett and S. Mendelson, "Rademacher and gaussian complexities: Risk bounds and structural results," *The Journal of Machine Learning Research*, vol. 3, pp. 463–482, 2003.
[37] T. Evgeniou and M. Pontil, "Regularized multi–task learning," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004, pp. 109–117.
[38] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Transactions on Consumer Electronics*, pp. 76–84, 2011.
[39] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research," in *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, vol. 25. Citeseer, 2011, pp. 59–62.
[40] K. Weekly, M. Jin, H. Zou, C. Hsu, A. Bayen, and C. Spanos, "Building-in-briefcase (bib)," *arXiv preprint arXiv:1409.1660*, 2014.
[41] M. Jin, R. Jia, and C. Spanos, "Apec: Auto planner for efficient configuration of indoor positioning systems," in *The 9th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM'15)*, 2015, pp. 100–107.
[42] X. Jiang, S. Dawson-Haggerty, J. Taneja, P. Dutta, and D. Culler, "Creating greener homes with ip-based wireless ac energy monitors," in *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, 2008, pp. 355–356.
[43] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart*: An open data set and tools for enabling research in sustainable homes," 2012.
[44] Y. Nesterov, "Gradient methods for minimizing composite functions," *Mathematical Programming*, vol. 140, no. 1, pp. 125–161, 2013.
[45] V. López, A. Fernández, J. G. Moreno-Torres, and F. Herrera, "Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. open problems on intrinsic data characteristics," *Expert Systems with Applications*, vol. 39, no. 7, pp. 6585–6608, 2012.
[46] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420–430, 2014.
[47] R. H. Dodier, G. P. Henze, D. K. Tiller, and X. Guo, "Building occupancy detection through sensor belief networks," *Energy and buildings*, vol. 38, no. 9, pp. 1033–1043, 2006.
[48] W. Kleiminger, C. Beckel, and S. Santini, "Opportunistic sensing for efficient energy usage in private households," in *Proceedings of the Smart Energy Strategies Conference 2011*, Zurich, Switzerland, 2011.
[49] O. Shih and A. Rowe, "Occupancy estimation using ultrasonic chirps," in *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*. ACM, 2015, pp. 149–158.
[50] S. V. Verdú, M. O. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *Power Systems, IEEE Transactions on*, vol. 21, no. 4, pp. 1672–1682, 2006.
[51] C. Beckel, L. Sadamori, and S. Santini, "Automatic socio-economic classification of households using electricity consumption data," in *Proceedings of the fourth international conference on Future energy systems*. ACM, 2013, pp. 75–86.
[52] D. De Silva, X. Yu, D. Alahakoon, and G. Holmes, "A data mining framework for electricity consumption analysis from meter data," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 3, pp. 399–407, 2011.
[53] M. Jin and C. Spanos, "Brief: Bayesian regression of infinite expert forecasters for single and multiple time series prediction," in *54th IEEE Conference on Decision and Control (CDC 2015)*, 2015.
[54] J. M. Abreu, F. C. Pereira, and P. Ferrão, "Using pattern recognition to identify habitual behavior in residential electricity consumption," *Energy and buildings*, vol. 49, pp. 479–487, 2012.
[55] S. Dawson-Haggerty, S. Lanzisera, J. Taneja, R. Brown, and D. Culler, "@scale: Insights from a large, long-lived appliance energy WSN," in *Proceedings of the 11th international conference on Information Processing in Sensor Networks*. ACM, 2012, pp. 37–48.
[56] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.



Ming Jin received the B.Eng. degree (honor) in Electronic and Computer Engineering from the Hong Kong University of Science and Technology. He is currently pursuing his Ph.D. degree in Electrical Engineering and Computer Science at the University of California, Berkeley. His current research focuses on the societal-scale cyber physical systems, from buildings to the (micro-)grid, by employing modeling, optimization, and control theories to design new algorithms and mechanisms. He received the best paper award in the 2015 International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies.



Ruoxi Jia received the B.S. degree in Electrical Engineering and Computer Science from Peking University, China. She is pursuing her Ph.D. degree in Electrical Engineering and Computer Science at the University of California, Berkeley. Her research interests include signal processing, statistical learning and information fusion with applications in smart buildings, in particular, indoor positioning, occupancy modeling, and building infrastructure fault diagnosis.

Costas J. Spanos has present research interests including the application of statistical analysis in the design and fabrication of integrated circuits, and the development and deployment of novel sensors and computer-aided techniques in semiconductor manufacturing. He is also working towards the deployment of statistical data mining techniques for energy efficiency applications, and is the Principal Investigator of the Singapore based SinBerBEST project, focusing on energy efficient buildings. In 2000 Prof. Spanos was elected Fellow of the Institute of Electrical and Electronic Engineers for contributions and leadership in semiconductor manufacturing, and in 2009 he was appointed in the Andrew S. Grove Distinguished Professorship, in the Department of Electrical Engineering and Computer Sciences.