A Survey on Intrusive Load Monitoring for Appliance Recognition

Antonio Ridi†*, Christophe Gisler†, Jean Hennebert†

*University of Applied Sciences Western Switzerland, College of Engineering and Architecture of Fribourg, IcoSys Institute {antonio.ridi, christophe.gisler, jean.hennebert}@hefr.ch
†University of Fribourg, Department of Informatics, Fribourg, Switzerland {antonio.ridi, christophe.gisler, jean.hennebert}@unifr.ch

Abstract—Electricity load monitoring of appliances has become an important task considering the recent economic and ecological trends. In this game, machine learning has an important part to play, allowing for energy consumption understanding, critical equipment monitoring and even human activity recognition. This paper provides a survey of current researches on Intrusive Load Monitoring (ILM) techniques. ILM relies on low-end electricity meter devices spread inside the habitations, as opposed to Non-Intrusive Load Monitoring (NILM) that relies on an unique point of measurement, the smart meter. Potential applications and principles of ILMs are presented and compared to NILM. A focus is also given on feature extraction and machine learning algorithms typically used for ILM applications.

I. INTRODUCTION

Nowadays, appliances are responsible of a significant part of the electricity bill in residential and commercial buildings. For instance in U.S. residential building, lighting and appliances represent 30% of the electricity consumption [1]. In this context, Appliance Load Monitoring (ALM) has become a key application for consumption understanding and energy savings. Two approaches are existing:

1) Non-Intrusive Load Monitoring - NILM consists in measuring the electricity consumption using a smart meter, typically placed at the meter panel. Relying on a single point of measure it is also called one-sensor metering. The qualification of non-intrusive means that no extra equipment is installed in the house. With NILM, the appliance signatures are superposed and, for comprehending the contribution of single appliances, they have to be separated. This operation is called disaggregation of the total electricity consumption.

2) Intrusive Load Monitoring - ILM consists in measuring the electricity consumption of one or few appliances using a low-end metering device. The term intrusive means that the meter is located in the habitation, typically close to the appliance that is monitored.

As illustrated in Fig. 1, there is actually a continuum between NILM and the most granular ILM where there would be one meter per equipment. For this reason we propose to define 3 sub-domains:

- ILM 1 Such systems rely on submeters that are measuring the consumption after the primary utility energy meter. Submeters are typically used to monitor a zone of the house and can be placed at the circuit breaker level.
- ILM 2 The meters are located at the plug level. The appliances connected to the outlet or multi-outlet are directly monitored.
- ILM 3 Single appliances are monitored with meters directly embedded in the equipment or in the outlet dedicated to the appliance.

Our review of the literature shows that most of the research goes now in NILM and ILM 3 domains.

Fig. 1. Distinction between ILM and NILM.

Relying on multiple points of measurement, ILM is also referred to as multi-sensor metering [2]. Other authors are also using equivalent denominations such as Intrusive Appliance Load Monitoring systems (IALMs) [3][4], Hardware-Based Sub-Metering [5][6] and Distributed System / Metering [7][8]. In this paper we will use ILM as generic term.

ILM can be used for numerous applications:

- Local energy consumption understanding. The objective is here to provide the household with energy feedback on single appliance. This feedback is typically given directly on the plug or on static or mobile displays [9]. Closely related, applications are also attempting to predict the electricity consumption of different appliances [10].
- Global energy consumption understanding. Related to the first one, another important application is to
allow a better understanding of the monthly electricity bill. The principle is to compute the relative contribution of each appliance to the global consumption through an aggregation of the different ILM sensors [11].

- **Appliance monitoring.** The objective is here to detect abnormal electricity consumption, deviation of consumption or faulty devices [12][13].

- **Evaluation of NILM environments.** NILM can also be employed for evaluating NILM disaggregation performances. In this case, the disaggregation algorithm is applied to the NILM smart meter data and the results are compared to the annotated ground truth provided by ILM sensor data [14][15].

- **Simulation of NILM environments.** Data acquired from several ILM sensors can also be artificially aggregated to simulate a NILM environment and perform evaluation of disaggregation algorithms as explained in the previous point [16].

- **Human activity recognition.** ILM has also demonstrated its potential for the indirect detection of human activities through electricity monitoring [17][18].

- **Appliance localization.** The consumption signature can be used to identify the appliance and the position of the ILM sensor provides an indication of its position [19][20].

Generally speaking, these ILM applications rely on data management, signal processing, statistics and machine learning technologies. An important task for most applications is the matching of incoming unseen signals against previously seen signals. Typically, models are created using historical “training” data and, according to the labels associated to these data, one can perform appliance identification, brand identification and state identification. In this paper we will provide more details on signal processing and machine learning approaches applied to these domains, pointing to the best-practices in this field.

In Section II, ILM and NILM approaches are compared in further details. In Section III, we describe the hardware architecture used to implement ILM systems. In Section IV and V, more details are given on feature extraction and machine learning techniques typically used in ILM systems. Conclusions and perspectives are finally presented.

II. NILM vs ILM

Scientific publications about NILM outnumber the publications about ILM. With publications starting in the 90’s, NILM is actually an elder research topic than ILM [21][22]. Datasets for NILM have also been made available for disaggregation tasks, contributing to its momentum [23][24][25]. NILM has also several advantages compared to ILM. Given its one-sensor based nature, installation is easier and data acquisition is simpler. ILM is typically based on cheaper meters, however its cost scales linearly with the number of sensors.

ILM presents also several advantages over NILM. First, finer information is acquired as the sensors are more numerous by definition. Finer details about appliance signatures are also made available and ease their modelling. Typically, low power appliances or appliances in stand-by are difficult to be identified with NILM [26], while their detection is feasible using ILM. Recently, stand-by power became one of the largest source of consumption in residences [27], representing up to 26% of the total energy consumption [28]. Another drawback of NILM is in the difficulty to detect devices with multiple consumption states or devices showing continuously variable energy use [26] (see also Section III).

In a similar way as what has been proposed for NILM in [22], we can distinguish two ILM cases for the identification task:

1) **MS-ILM** - manual setup ILM refers to the training of models on data acquired in the environment. A manual intervention of the user is necessary to label “on-site” the data according to the appliance. The appliance recognition performances are good as the signals are emitted by the same appliance that were used to train the models. In this case, only temporal variability is impacting the performance of the models.

2) **AS-ILM** - automatic setup ILM do not require the acquisition and labelling of on-site data. The system is a priori trained and may receive signals from “unseen” appliances. Not only temporal variability but also intra-class variability may impact the recognition performance. Intra-class variability is here due to difference of brands or models in the same class.

On the one hand, MS-ILM is more expensive than AS-ILM in terms of setup time. On the other hand, MS-ILM is more precise than AS-ILM [6][29][30][31]. Similar conclusions were also reported for NILM in [32].

III. APPLIANCE MONITORING

Limiting ourselves to the residential sector, we enumerated more than 50 different categories of appliances. Different brands or models from the same category may also show significant differences in terms of electricity consumption. Considering machine learning, the training of models for each category could then become a rather complex task requiring large signature databases. Further to this, appliances can be in different states according to their use or functioning. Up to four states can be listed as explained in [22] and updated in [33][34]:

- **Type I - on/off** with only two states either on or off, such as light bulbs or toasters.
- **Type II - finite states** corresponding to different modes of energy consumption, such as microwaves, fridges and tv sets.
- **Type III - continuously variable** as a function of time, typically for battery charging appliances such as laptops or phones.
- **Type IV - permanently on** with a constant source of consumption, as alarms or landline phones.
These types actually emphasize the need for a mix of different techniques for signal processing and feature extraction, as further described in Section V.

In Table 1, we illustrate the diversity of coverage of appliance categories used for identification tasks in different research work. For sake of simplicity, we limited to categories with more than 4 occurrences. This Table clearly shows a large variability, making it difficult to compare the different research works in terms of performance.

### TABLE 1. APPLIANCE CATEGORIES USED IN RESEARCH WORKS.

<table>
<thead>
<tr>
<th>Sensors</th>
<th>Gateway 1</th>
<th>Gateway 2</th>
<th>Gateway 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILM1 sensors</td>
<td>ILM2 sensors</td>
<td>ILM3 sensors</td>
<td></td>
</tr>
<tr>
<td>Drivers to sensors/gateways</td>
<td>Data processing applications</td>
<td>Historical Data</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. ILM System Architecture

![ILM System Architecture](image)

TABLE I. APPLIANCE CATEGORIES USED IN RESEARCH WORKS.

<table>
<thead>
<tr>
<th>Appliance</th>
<th>[35]</th>
<th>[36]</th>
<th>[37]</th>
<th>[38]</th>
<th>[39]</th>
<th>[40]</th>
<th>[41]</th>
<th>[42]</th>
<th>[43]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fridge</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Microwave</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Wash machine</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Coffee maker</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Computer</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Laptop</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Printer</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Monitor</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Television</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Lamp</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Vacuum cleaner</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Fan</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Iron</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Hair dryer</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Phone charger</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

IV. SYSTEM ARCHITECTURE

A typical ILM system architecture is illustrated in Figure 2. The bottom layer is the network of sensors that may fall in the previously defined categories ILM1 (sub-meters), ILM2 (plug level) or ILM3 (appliance level). Sensors can also behave as actuators with the ability to switch on or off appliances. The next layer is composed of gateways that relay the information from the sensor up to the server. Gateways contains drivers and protocols enabling the communication with the sensors over potentially different networks such as PLC, IP, EnOcean or EIB/KNX [18][43]. The gateway layer is actually optional in the case, for example, of IP based sensors with a direct connection to the server. The server layer is in charge of communicating with the gateways and/or sensors, storing the data, processing the data and providing the necessary information to render vues in the upper layer. Examples of ILM systems have been described for appliance monitoring [13], appliance identification [41][29][7] and consumption prediction [18].

The weakness of such centralized architecture is in the server that is a single point of failure. Other types of systems will probably emerge with more distributed architectures relying on memory and computing capacities embedded in the sensors, typically based on Internet of Things paradigms [44].

V. DATA ACQUISITION AND FEATURE EXTRACTION

Depending on the target application, the sampling frequency of the electricity consumption may vary. In some works, a rather high sampling frequency from 1 kHz up to almost 100 kHz is used [29][42]. In this case, the identification results are more precise, typically allowing to capture finer state transitions and eventually separating brands in the same category. As a drawback, the equipment of such systems is more costly [45]. Other works have shown the possibility to use cheaper meters with limited sampling frequencies, typically less than 10 Hz [39][31][6].

A. Time-dependent features

As in the NILM, common temporal features are extracted and used in ILM such as the real (active) power (P), the reactive power (Q). Some appliances can be easily distinguished in the P-Q space depending on their resistive, capacitive and inductive characteristics [46]. From these features, others can be computed, as the complex power and the apparent power. The voltage (V) and current (I) are as well very popular as features [29]. The voltage-current (V-I) trajectories have also been proposed for characterizing the appliances [37]. The root mean square and the peak values of the current are also used, for example in [41]. Information about the current form, as the crest factor [42], the form factor or their combination have been used in [41]. The inrush current compared to the steady state current was also used in [8].

Pre-processing operations such as normalization and phase shifting are sometimes applied [40]. In some cases, a reduction of the feature space is also applied using Principal Component Analysis (PCA) [29].

Features based on the signal evolution, as the first and second derivative, have been proposed in [39][31]. Such features have shown to bring useful information in the classification task. Other features are based on counting the number of occurrences of events in a period of time, such as the number transition between power intervals in [6], the number of edges in [35] and the number of threshold crossing in [36]. Other features can be added to the list, as the maximum, minimum or average of power in a certain interval of time [6][36].
B. Frequency-based features

Frequency analysis is also used in the context of appliance recognition, especially when the acquisition frequency is medium to high (from 1 kHz up to 100 kHz). Discrete Fourier Transformation (DFT) or Fast Fourier Transformation (FFT) are typically applied. DFT is reported to be less efficient when the sampling frequency is low [35]. FFT is used in [36] to determine the highest frequency in the signature and the energy contained in a the band of 5%, 15% and 25% of the highest frequency. The impact of FFT-based features has been evaluated showing differences in performances of the system when injecting other features derived from the current in time [8]. Using FFT on signals with a sampling frequency of 20 kHz, appliance harmonic content is unique and even devices of the same category can be distinguished [47]. Other authors reported detailed analysis of harmonics using high sampling frequency [42].

VI. MACHINE LEARNING

Machine learning implies the use of historical data to build models. Many research works report on the use of data recorded through acquisition campaigns using commercial sensors or specific hardware such as in [48][49][12]. As alternative, publicly available databases can be used. Dedicated to appliance recognition, the Tracebase\(^1\) database contains more than a thousand of electrical appliance signatures, recorded from 122 appliances spread into 31 categories [36]. The ACS-F1\(^2\) database contains 200 appliances consumption signatures recorded from 100 appliances of different brands and models spread into 10 categories [30]. The ACS-F1 is smaller than the Tracebase database but presents the advantage of being balanced in terms of categories. As well, two evaluation protocols are defined with the ACS-F1 database, roughly corresponding to MS-ILM and AS-ILM scenarios presented in Section II. The signatures in both databases are recorded at low frequency. Other databases have been made available but are rather dedicated to statistical analysis [50][51].

Most of the machine learning approaches are based on supervised techniques in the context of appliance identification. Paradiso et al. [6] are proposing the use of multilayer perceptrons trained with back-propagation. Low sampling frequency data are used from which power-based features are extracted such as the number of samples and transitions between power intervals. Tests are reported for MS-ILM and AS-ILM systems. Accuracy of 95.26% is reported for MS-ILM on 8 monitored devices. Kato et al. [29] propose to use one-class Support Vector Machines (SVM) with Gaussian kernels. The features are extracted by voltage and current wave-shape using a sampling frequency of 6 kHz. In the MS-ILM case they used 16 appliances achieving 99.9% of accuracy rate. In the AS-ILM case they used 25 appliances achieving 95.8% accuracy. Zufferey et al. [38] compare K-Nearest Neighbors (k-NN) and Gaussian Mixture Models (GMMs) using 6 appliance categories. The electric measurements are sampled at 10\(^{-1}\) Hz using low-end Plogg meters and the features are power-based. Their best reported performance is 85% accuracy using k-NN.

A continuation of this work is presented by Ridi et al. [39][31]. They report about MS-ILM and AS-ILM systems using the ACS-F1 database previously described [30]. In both cases, they compared K-NN and GMMs. The feature space is also extended including temporal dynamic coefficients. For MS-ILM, 93.8% accuracy is reported using GMMs on 10 categories. For the more difficult AS-ILM task, 74% accuracy is reported with GMMs on 10 categories. Reinhardt et al. report on several works [36][8][52], comparing various algorithms such as Bagging, Bayesian Network, J48, JRip, LogitBoost, Naive Bayes, Random Committee, Random Forest and Random Tree for the classification of appliance categories using a MS-ILM scenario. In a first work they use data from the Tracebase database recorded at low frequency. The best reported performance is up to 95.5% accuracy using Random Committee. In another work, the same classification algorithms are used on higher frequency data (1.6 kHz) from which current-based and harmonics features are extracted. Better performances are reported up to 100% for Bayesian network. Englert et al. [42] present an hardware and software system treating high frequency data (96 kHz). Temporal and frequency features are computed from the voltage and current values. Different tasks are evaluated including category recognition, equipment identification and operation mode recognition. For category recognition, 99.8% accuracy is reported using Random Forest classifiers on 14 categories. In the work of F. K. Adeel Abbas Zaidi and P. Palensky [35], Dynamic Time Warping (DTW) and Hidden Markov Models (HMMs) are proposed. The signatures are sampled at 10\(^{-1}\) Hz and several features are extracted and benchmarked. HMMs are reported to perform better than DTW using a 6 category task. Saitoh et al. [40] reports on the use of k-NN for the identification of 35 appliances in a MS-ILM scenario. The signatures are sampled at 4.4 kHz from which current-based features are extracted. Accuracy is reported for category recognition, appliance recognition and state recognition, with, respectively 85.5%, 80.5% and 76.3%. Fitta et al. [53] propose to recognize the switch-on and off events from appliance signatures acquired at 6.4 kHz. They used K-NN systems trained on extracted features such as the real power and current harmonic. Performance on this 2-class task of 87.5% and 90.6% correct identification is reported for, respectively switch-on and switch-off events. Other works reporting rather few algorithmic details on ILM tasks can be found in [7][41].

Rather few unsupervised techniques are reported for ILM tasks. Lam et al. are proposing the use of Dendrograms for appliance representation and clustering [54][37]. Shape features of trajectories in the V-I space are used as features. The Dendrograms allowed to cluster 30 appliances into 13 groups showing similar operating characteristics.

In this survey, we observed that machine learning algorithms are well spread in papers without a preferred technique. The appliances and features choice have consequences on the pattern separability in the feature space and therefore influence the selection of machine learning algorithms.

VII. CONCLUSION

We observe an increasing number of papers published over the last 3 years about ILM tasks, showing an increasing interest
of the scientific community. The survey of about 50 papers related to load monitoring and more specifically to ILM allows us to draw the following conclusions:

1) **Data availability.** ILM currently lacks of large databases that are publicly available, especially if compared to NILM. Comparison between scientific work is currently made difficult as many teams work on their own private data sets using specific protocols. Initiatives such as Tracebase or ACS-FI will potentially foster the scientific community towards comparable evaluations.

2) **Appliance sets.** The type of appliances and their occurrences in the data sets are probably influencing the reported performances in a significant manner. In general, appliances having similar resistive, capacitive and inductive characteristics are more difficult to separate in the feature space. This fact leads to two consequences. First, one should analyze the types of appliances presented in scientific papers to interpret fairly the reported performances. However, such information is sometimes not available in papers. Second, current reported results are probably optimistic as larger data sets with more categories will show lower performances.

3) **Sampling frequency.** In general, we observe that higher sampling frequency of signatures give better classification results. More detailed signal information is indeed available, allowing to use frequency-based features that reveal especially powerful for appliance identification inside the same category. However, medium-high frequency metering equipments are more expensive than low frequency sampling meters.

4) **Machine learning.** There is currently no "leading" or preferred machine learning technique. This is probably due to the variability of the sets of appliance categories, and due to the different types of measurements, sampling frequencies and feature selections.

VIII. ACKNOWLEDGMENT

This work is supported by the grant Smart Living Green-Mod from the Hasler Foundation in Switzerland and by the HES-SO.

**REFERENCES**


