A Comparison of Generative and Discriminative Appliance Recognition Models for Load Monitoring

Ahmed Zoha [#1], **Muhammad Ali Imran** [1], **Alexander Gluhak** [1], **Michele Nati** [1]

[1]Centre for Communication Systems Research, University of Surrey, GU2 7XH, Guildford, UK

E-mail: `a.zoha@surrey.ac.uk`

Abstract. Appliance-level Load Monitoring (ALM) is essential not only to optimize energy utilization but also to promote energy awareness amongst consumers through real-time feedback mechanisms. Non-intrusive load monitoring is an attractive method to perform ALM that allows tracking of appliance states within the aggregated power measurements. It makes use of generative and discriminative machine learning models to perform load identification. However, particularly for low-power appliances, these algorithms achieve sub-optimal performance in a real world environment due to ambiguous overlapping of appliance power features. In our work, we report a performance comparison of generative and discriminative Appliance Recognition (AR) models for binary and multi-state appliance operations. Furthermore, it has been shown through experimental evaluations that a significant performance improvement in AR can be achieved if we make use of acoustic information generated as a by-product of appliance activity. We demonstrate that our a discriminative model FF-AR trained using a hybrid feature set which is a catenation of audio and power features improves the multi-state AR accuracy up to 10 %, in comparison to a generative FHMM-AR model.

1. Introduction

It is well known that due to growing energy demands, specifically in the last few decades, energy conservation is becoming challenging. This is particularly true for residential spaces and within offices that accounts for almost *40* % of total European Union (EU) energy consumption, generating almost *36* % of Green House Gas (GHG) emission. Moreover, it has been predicted that global energy demands will double by the end of 2030 [1], which leads to increased energy prices along with its negative implications (i.e., CO_2 emissions) on the environment. This substantial growth in energy consumption directly impacts the economy of a country. In this context, information and communication technologies (ICT), more specifically Internet of Things (IoT), can prove to be important enablers of smart solutions that reduce the overall energy consumption. A detailed review [2] of more than 60 feedback studies suggest that maximum energy saving can be achieved using direct feedback mechanisms (i.e., real-time appliance level consumption information) as opposed to indirect feedback mechanisms (i.e., monthly bills, weekly advice on energy usage). To achieve this real-time energy information we see a large scale deployment of smart meters by the governments of UK and USA. However, these meters provide little information on the breakdown of energy spent. This has motivated research efforts to develop Appliance-level Load Monitoring (ALM) methods to achieve fine-grained energy
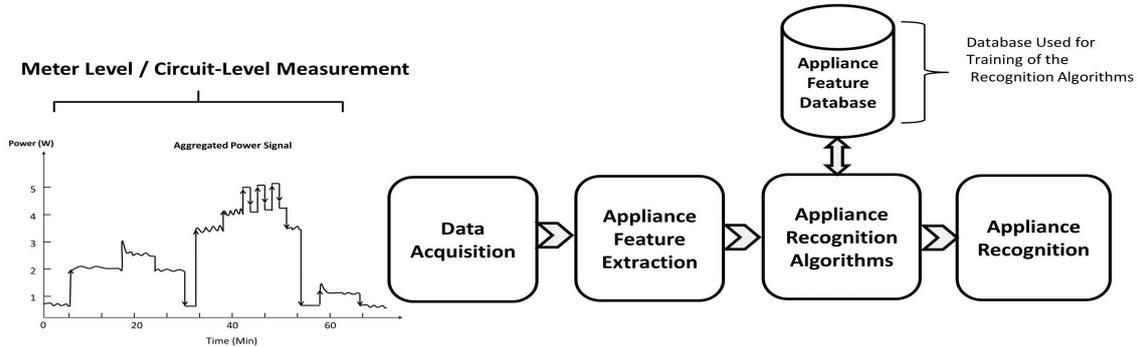
Figure 1. A General Framework for Non-Intrusive Load Monitoring

monitoring. The goal of the detailed energy information acquired from these ALM methods is to make the invisible flow of energy visible to the consumers. This has a direct impact on the behavior of the users, as mostly they are unaware of their energy wastage behavior. Moreover, energy-awareness is important so that users can differentiate between different energy providers, and develop their own strategies for energy saving. Consumers are able to cut down their power usage through knowing what devices are often used purposelessly. From a smart-grid perspective, demand side management could be supplemented by having an analysis of appliance level energy usage information which provides a better insight into power utilization and facilitates formation of optimum resource allocation strategies. ALM is not only a pre-requisite for precise energy feedback, but it is equally applicable for other applications such as fault detection, monitoring and security as well as for the automated energy management systems.

The remainder of the paper is organized as follows. In the next section, we review related research work, highlighting the limitation of current ALM approaches. We present an overview of our generative and discriminative AR models in Section 3. The experimental setup and performance comparison of our proposed models have been reported in Section 4. Finally, we provide conclusion and future work in Section 5.

2. Load Monitoring Approaches

Research efforts to achieve finer granularity of energy-consumption statistics has led to the development of intrusive and Non-intrusive load monitoring (NILM) techniques. A major drawback of intrusive load monitoring system is that, in order to measure appliance-level energy consumption, multiple sensors needs to be installed on each target appliance. Though we can achieve a accurate measure of device energy usage through this approach but high installation complexity, cost, as well as sensor calibration and data aggregation are outstanding issues that will not favor the use of this technique. NILM, on the other hand uses aggregated power measurements acquired from a single point (i.e., meter or circuit-level) and employ feature extraction methods to extract device features. These features are further used to model AR learning algorithms to disaggregate device specific power usage patterns. Lately, NILM methods have gained a lot of attention in the research community, because of its ability to overcome some of the major challenges faced by intrusive load monitoring systems, as highlighted above.

A general framework for NILM is shown in Figure (1), which involves feature extraction followed by an application of AR algorithm. A change in the aggregated power measurements is observed whenever an appliance switches its state (i.e., ON to OFF), which can be further characterized into steady state and transient events, as proposed by Hart [3]. Accordingly, steady-state and transient-state feature extraction methods are being developed to create a

unique signature that will identify each appliance. These features are monitored for tracking load operations in the power measurements and further classified to a particular load using intelligent classification algorithms. The classification algorithms need to be trained by providing labeled examples for training, which first requires extraction of device signatures independently to form an appliance feature database. The AR algorithms reported in literature are based on either pattern recognition [4, 5] or optimization based approaches [6, 7]. Comparatively, pattern recognition approaches that are based on either discriminative or generative machine learning models have shown to perform better than optimization approaches [8]. However, load identification is highly dependent on extracted appliance signatures particularly for discriminative AR models.

The choice of using either steady-state or transient signatures has its own advantages and disadvantages. Steady-state load signatures represent the steady-state behavior of the appliance; therefore it can be obtained from the low-frequency sampling of current and voltage waveforms. High-power appliances exhibit distinct steady-state signatures, however devices with low-power consumption profile are difficult to disaggregate/recognize from the aggregated load measurements due to overlapping steady-state features. In contrast, transient signatures require high sampling of power waveforms to extract shape, size, duration and high order harmonics features to characterize an appliance state transition event. Although, transient features in conjunction with steady-state features have shown to improve the overall disaggregation accuracy of AR models, it requires costly hardware, in addition they are sensitive to the wiring architecture and demand excessive training of the algorithms. Moreover, if the system needs to be envisioned for residential spaces, current smart meters are only able to provide data at a low frequency resolution. Nevertheless, most of the existing solutions reported in literature [8] achieve sub-optimal performance in a real-world environment due to following reasons:

- Firstly, most research work is focused on identifying HVAC appliances whereas the presence of low-power appliance loads is often neglected. It is still a challenge for traditional NILM solutions to identify low-power appliances with high accuracy. In our work, we show that using circuit-level power measurements it is possible to accurately discern low-power appliances which are particularly of interest to profile energy consumption of the users within offices.

- Secondly, it is often the case that only binary state (i.e., ON or OFF) operation of the appliances are considered for evaluating the performance of AR algorithms. However, in a real-world setting, appliances often operate in multiple states. Therefore, we have considered binary as well as multi-state operation of the devices in our experimental evaluations, along with adequate feature set selection for optimal classification.

- Thirdly, as highlighted above, AR models trained on steady-state features alone achieve low recognition accuracy, particularly for low-power appliances due to similarity of the appliance features. We propose a discriminative AR model that make use of auxiliary acoustic information associated with the appliance operational state, in order to overcome the ambiguous steady-state feature overlap. Moreover, a comparison has been provided with a generative AR model that employs steady-state features alone.

In the next section, we provide an overview of our discriminative and generative AR models.

3. Our Proposed AR Models
3.1. Factorial Hidden Markov Model Based Appliance Recognition Model (FHMM-AR)
In [9] we propose a generative AR model based on Factorial Hidden Markov Model (FHMM-AR) in order to perform load monitoring in the aggregated power measurements. The FHMM-AR tries to model the joint distribution of hidden appliance states and the power measurements. A structured variational approximation method is adopted for learning, whereas inference is

Table 1. Features used to train our proposed AR models [*]

Features	Description
Features for Generative Model f_G	$f_G = [P,\ Q,\ \Delta P,\ PF, P_{std}]$, where P and Q are the active and reactive power, PF is the power factor, and P_{std} is the standard deviation of power values within each frame
Audio Features f_A	13 coefficients of Frequency Filtered Band Energies (FFBE's) obtained by applying a second order filter $H(z) = z - z^{-1}$ at the output of mel-scaled filter bands
Features for Discriminative Model f_D	$f_D = \{f_A + f_{Dp}\}$ whereas f_{Dp} is the catenation of $[P,\ Q,\ I_{RMS},\ V_{RMS},\ PhaseAngle]$

[*] We refer the reader to [9] and [10] for a detailed description of the feature sets mentioned in the Table 1

performed using standard Baum-Welch procedure. The feature vector f_G as shown in Table (1) is selected after performing optimal feature selection, and is used to train the final FHMM-AR model. The complete algorithmic details of FHMM and results of feature selection evaluations have been provided in [9].

3.2. Feature Fused Appliance Recognition Model (FF-AR)
In [10] we propose a discriminative AR model based on Support Vector Machines (SVM), whereas the selection of optimal feature set and comparison with other machine learning models have also been presented. It has been shown that proposed model when trained using steady-state features alone achieves low recognition accuracy, due to their ambiguous overlapping in the feature space. To overcome this problem, a solution is proposed that exploits the complimentary acoustic information which is generated as a byproduct of appliance operation. A number of audio features have been extracted and optimal feature selection is performed along with the power features to jointly characterize the audio and power events. It has been demonstrated that the same discriminative model trained using a fused feature vector which is a catenation of f_A and f_{Dp} as shown in Table (1) improves the discriminative ability of classifier. In the text to follow, we refer this model as FF-AR model which has been compared against the FHMM-AR model in this paper, for recognizing binary and multi-state operation of the appliances.

4. Experimental Evaluation
4.1. Experimental Setup
The experimental setup as shown in Figure (2) is similar to the one reported in [9]. To train and test the appliance models, data is gathered from a target user desk via a smart power outlet. This smart outlet has a inbuilt energy meter co-located at user work desk so that 5 target appliances including Work Station (WS), LCD, Laptop, Desk Lamp and a fan can be attached to it via multi-socket. In addition, it has a audio sensor interfaced with it to monitor the acoustic activity of the environment, so that appliance acoustics (i.e. Keyboard typing sound and Fan Operational Sound) could also be captured and processed. This smart power outlet acts as a circuit-level monitoring device which logs aggregated power readings as well as audio measurements of all the appliances at $3Hz$ and $16KHz$, respectively. The acquired data is transmitted to a selected aggregation point (sink), whereas the sink is further connected to a Management Gateway (GW) which reports the data to a central server, so that it can be stored in a database. The monitoring station queries the database periodically to acquire multi-modal measurements for off-line processing.
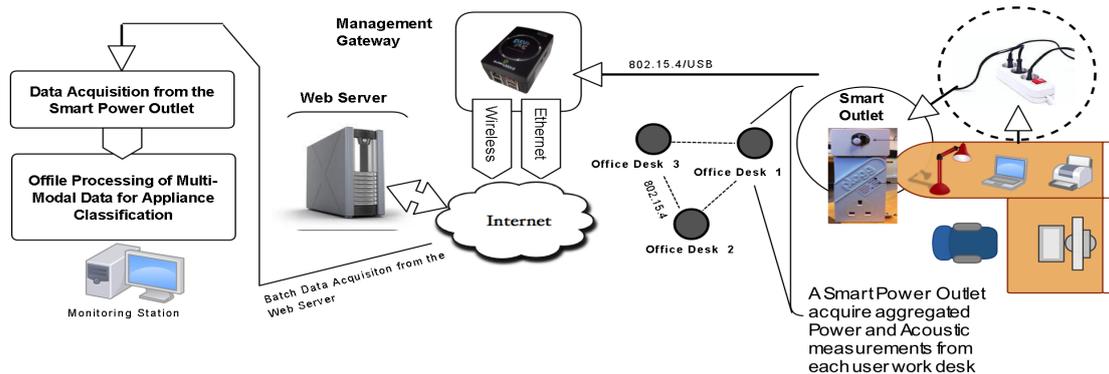
Figure 2. Experimental Test Bed Architecture for Appliance Monitoring at user work desk

Table 2. Disaggregation Accuracy of FHMM based AR Model versus Feature Fused AR Model

Test No	Appliance Combinations	FHMM-AR		FF-AR	
		Binary	Multi-State	Binary	Multi-State
1	WS, LCD	0.987	0.956	0.960	0.950
2	Laptop,Lamp	0.980	0.961	0.950	0.930
3	Lamp, Fan	0.976	0.87	0.940	0.920
4	Laptop, Fan	0.862	0.702	0.943	0.90
5	WS, LCD,Lamp	0.941	0.890	0.950	0.923
6	WS, Lamp,Laptop	0.931	0.91	0.930	0.910
7	WS, LCD,Lamp,Laptop	0.90	0.74	0.921	0.890
8	WS, LCD,Lamp,Fan	0.88	0.731	0.911	0.870
9	WS, LCD,Laptop,Fan	0.839	0.669	0.892	0.862
10	**All 5 Appliances**	**0.768**	**0.614**	**0.879**	**0.842**

To evaluate the performance of our proposed models, the experiments are conducted in two phases: Binary and Multi-state operational phase. In the Binary phase, all the appliances are forced to operate only in their ON or OFF state disabling any intermediate state transitions (i.e., idle or standby). Conversely, for multi-state operations all possible state transitions have been taken into account for all target appliances. To train FHMM-AR and FF-AR models, separate data for each appliance has been collected during their binary and multi-state operations. A total of 10 test cases has been designed, each representing a unique set of appliance combination. An average duration of 30 minutes having approximately 120 events per test case have been generated by manually changing the respective states of the appliances. A 10 fold cross validation strategy has been adopted to report the performance of the models in terms of F-measure. In the subsequent section, we will summarize our findings from our experimental evaluations.

4.2. Results

The result for 10 different appliance combination scenarios has been summarized in Table (2). It has already been highlighted in our previous research work [9] that the performance of generative model FHMM-AR severely degrades under the following conditions

- Firstly, as the number of appliance states starts to grow the possibility of similar power draw values between appliance intermediate states also increases. This impacts the overall classification accuracy of the FHMM-AR models, as it can be seen from the results reported

for test cases 5 to 10.

- Secondly, not only the number, but the type of appliances operating in parallel also has an impact performance of the FHMM-AR models. This can be seen from the results for all the test cases reported in Table (2), which contains a combination of fan, laptop and LCD. It is because FHMM-AR models makes an assumption during the model generation that the mean μ_t of respective appliance states linearly combine at the output as discussed in [9]. However, this is not true for cases when inductive and capacitive elements operate in parallel, such as fan and a laptop. Due to the leading and lagging power factor, the reactive powers of the load cancels each other instead of addition, thus violating the assumption. This is why, we see a low recognition performance of our generative model specifically for test cases $4, 8, 9$ and 10.

On the other hand, our discriminative model FF-AR perform significantly better than FHMM-AR models, as it can be seen from the results shown in Table (2). This is mainly because the classification accuracy of the discriminative models depends mainly on discriminative ability of the feature sets used to train the appliance models. FF-AR model is trained using a hybrid feature vector that increases the separability of appliance classes in the feature space, which consequently improves the AR accuracy. The keyboard typing sound of a laptop and the operational sound of a table fan, when jointly characterized with the power features using feature catenation helps removes the ambiguous overlapping between the two target classes. Therefore, we see a much higher F-measure for FF-AR model for test cases $4, 8, 9$ and 10 in comparison to FHMM-AR model as shown in Table (2). Although, we notice a performance degradation for the FF-AR model as the number of operating appliances increases, however if we look at test case 10 the FF-AR model achieves almost 11 % and 22 % improvement over FHMM-AR model for the case of binary and multi-state load classification, respectively. The overall average improvement achieved by the FF-AR model for all the test cases is about 3 % for binary and 10 % for multi-state device operations. This clearly indicates that unlike FHMM-AR models, the performance of the FF-AR models still maintain an acceptable performance even if the number of target appliances increases, however at the cost of increased complexity. We believe that although the use of additional audio modality increases the complexity of the system, however it is critical for the real-world systems to maintain an acceptable performance under all circumstances. Moreover, unlike other existing intrusive load monitoring systems reported in literature, our proposed multi-modal system as discussed in [10] maintains its non-intrusiveness by just employing only one audio sensor for the acoustic surveillance of target environment.

5. Conclusion and Future Work

This paper provides a comparison of generative and discriminative AR models in order to recognize the binary and multi-state operation of low-power appliances. We have demonstrated that our proposed discriminative model FF-AR is more suited for load recognition in comparison to our generative FHMM-AR model. Unlike FHMM-AR, the performance of FF-AR is less effected by the type and the number of target appliances operating simultaneously. The fusion of audio and power feature vectors clearly aids the classification process, resulting in a significant performance improvement for the FF-AR model. Our initial investigations demonstrate that in order to achieve acceptable AR performance in a real-world setting, it is promising to combine acoustic modality as a complimentary source of information, alongside power sensing. In future, we plan to extend this work by including more appliances in the dataset and further modify our algorithms for on-line load classification instead of performing off-line inference.

References

[1] Uteley J and Shorrock L 2008 Domestic energy fact file 2008 Tech. rep. Building Research Establishment, Garston UK

[2] Ehrhardt-Martinez K, Donnelly K A and Laitner J A 2010 Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities Tech. Rep. E105 American Council for an Energy-Efficient Economy (ACEE) Washington, DC

[3] Hart G W 1992 *In IEEE Proc.* **80** 1870–1891

[4] Zeifman M and Roth K 2011 *In IEEE Trans. Consum. Electron.* **57** 76–84

[5] Farinaccio L and Zmeureanu R 1999 *In Energ. Build.* **30** 245–259

[6] Liang J, Ng S K K, Kendall G and Cheng J W M 2010 *In IEEE Trans. Power Del.* **25** 551–560

[7] Suzuki K, Inagaki S, Suzuki T, Nakamura H and Ito K 2008 Nonintrusive appliance load monitoring based on integer programming *In Proceedings of SICE Annual Conference, Tokyo, Japan* vol 174 pp 2742–2747

[8] Zoha A, Gluhak A, Imran M A and Rajasegarar S 2012 *Sensors* **12** 16838–16866

[9] Zoha A, Gluhak A, Nati M and Imran M A 2013 Low-power appliance monitoring using factorial hidden markov models *In 2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing* pp 527–532

[10] Zoha A, Gluhak A, Nati M, Imran M and Rajasegarar S 2012 Acoustic and device feature fusion for load recognition *In Proceedings of 6th IEEE International Conference on Intelligent Systems (IS'12),Sofia, Bulgaria* vol 1 pp 386 –392