Appliance Recognition with Combined Single- and Multi-label Approaches

Marco Manolo Manca
marcom.manca@unica.it
University of Cagliari
Cagliari, Italy

Anthony Faustine
anthony.faustine@iti.larsys.pt
ITI, LARSyS, Técnico Lisboa
Lisbon, Portugal

Lucas Pereira
lucas.pereira@tecnico.ulisboa.pt
ITI, LARSyS, Técnico Lisboa
Lisbon, Portugal

ABSTRACT

The problem of appliance recognition is one of the most relevant issues in the field of Non-Intrusive-Load-Monitoring; its importance has led, in recent years, to the development of innovative techniques to try to solve it. The use of methods such as V-I trajectory, Fryze Theory Decomposition and Weighted Recurrence Graph have proved effective in recognising both single (*Single Label*) and multiple active appliances (*Multi Label*). This paper presents a new way of approaching the problem by unifying *Single Label* and *Multi Label* learning paradigms. The proposed approach exploits feature extraction techniques which allow the detection of both activated/deactivated appliances and all active appliances given aggregate current signal. We evaluate the proposed approach on a PLAID dataset. The obtained results indicate combining single-label and mult-label learning strategies for appliance recognition provides improved classification results with an F-score of 0.91.

CCS CONCEPTS

• **Computing methodologies → Machine learning**.

KEYWORDS

NILM, Appliance Recognition, Single-Label, Multi-Label

ACM Reference Format:

Marco Manolo Manca, Anthony Faustine, and Lucas Pereira. 2022. Appliance Recognition with Combined Single- and Multi-label Approaches. In *6th International Workshop on Non-Intrusive Load Monitoring (NILM '22), November 9–10, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3563357.3566153

1 INTRODUCTION

One of the main sub-tasks of Non-Intrusive Load Monitoring (NILM) is appliance recognition, its aim is to recognize which appliance is activated by analysing energy measurements and classifying appliances through machine learning techniques. This problem can be approached from two perspectives: in the *Single-Label* approach, the goal is to recognize appliances using the individual device's on/off signals; on the other hand, in the *Multi-Label* approach, the aggregate signal is used to understand which devices are activated and which are not. [1]-[13]-[10]

A major challenge is to solve the two problems simultaneously; many works offer interesting solutions by addressing these with innovative feature extraction and classification techniques, but in a separate way. It has been shown how using Weighted Recurrence Graphs can help increase the performance of classification algorithms about appliance recognition [2], as well as using preprocessing techniques based on Fryze theory decomposition [4].

In order to provide a more accurate classification, deep neural networks are increasingly being used in the NILM [5]-[7]; their effectiveness is, however, undoubtedly constrained by the availability of rich data. For this reason, high-frequency signal data are often used, allowing for more efficient training of this network. [6]-[15]

The main focus of this work is to open up a new challenge, namely to tackle the two problems mentioned above using a single algorithm capable of combining the two tasks in such a way that the information sharing of the two models can increase performance. So far, a combined algorithm has been created and tested that would homogeneously share information from the two models. Still, it would be of great interest to create one that would allow the exchange of information between the two in an optimized manner. The proposed method was tested on the publicly available Plug-Load Appliance Identification Dataset (PLAID) [8].

2 METHODOLOGY

The proposed work aims to detect which appliance m has been switched on or off and simultaneously estimate the state $s_m(t)$ of all M appliances from the aggregate current and voltage signal $x(t)$, for $m = 1, \ldots, M$. The aggregate signal can be described by Equation (1):

$$x(t) = \sum_{m=1}^{M} y_m(t) \cdot s_m(t) + \sigma(t) \tag{1}$$

where $y_m(t)$ represents the individual appliance measurements and $\sigma(\cdot)$ the noise contribution.

2.1 Problem formulation

The problem can be formulated as follows: let $\mathbf{A} \in \mathbb{R}^{T \times d_1}$ be the set of the input features obtained from the measurements of the switched-on or switched-off appliance, computed from the aggregate signal like the difference between the before and after the event; let $\mathbf{X} \in \mathbb{R}^{T \times d_2}$ be the aggregated signal values after an event; $\mathbf{M} = \{1, \ldots, M\}$ the set of indices representing each appliance; $\mathbf{Y} \in \mathbb{R}^{T \times M}$ the set of each appliance measurement composing the aggregate signal and $\mathbf{s}(t) = [s_1(t), \ldots, s_M(t)]$ the vector containing the state of each appliance at time $t \in \{1, \ldots, T\}$, where

NILM '22, November 9–10, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9890-9/22/11…$15.00
https://doi.org/10.1145/3563357.3566153

$s_j(t) \in \{0, 1\}$. Then, given a dataset $\mathbf{D} = \{\mathbf{a}(t), m, \mathbf{x}(t), \mathbf{s}(t) | \mathbf{a}(t) \in \mathbf{A}, m \in \mathbf{M}, \mathbf{x}(t) \in \mathbf{X}, t = 1, \ldots, T\}$, the aim is to train a classifier that can predict at the same time t which appliance m is activated and the state vector $\mathbf{s}(t)$ using $\mathbf{a}(t)$, $\mathbf{x}(t)$. It is important to note that whenever at instant t^* one of the m devices is switched on or off, the current and voltage consumption are added or subtracted to the aggregate consumption $x(t^*)$. Hence, this work assumes that appropriate information sharing during the training phase of both models will allow each task to be solved more efficiently. We will refer to the first task as *Single Label* problem and the second one as *Multi Label* problem.

2.2 Pre-processing and Feature Extraction

In this work, features derived from the current and the voltage of switching appliances on/off were used in the *Single Label* case; features derived from the aggregate current and aggregate voltage were used in the *Multi Label* case. In both cases, these are high-frequency measurements and one-cycle steady-state current $i(t)$ and voltage $v(t)$ signals were used, extracted as in [2], [3].

With the application of Fryze Theory Decomposition, which has been shown to improve performances because it can provide a distinctive feature for classification [14], these features are subsequently used to calculate the decomposed current features, i.e. given the signals $i(t)$ and $v(t)$, it is possible to decompose $i(t)$ into two components via Equation (2):

$$i(t) = i_a(t) + i_f(t) \tag{2}$$

where $i_a(t)$ and $i_f(t)$ represent respectively the active and non-active components of the current.
Once this is done, the Piece-wise Aggregate Approximation is used to reduce the dimensionality of the feature to a predetermined w, set to 50 in our work.

To the current signals thus obtained the Euclidian distance function is then applied:

$$d_{jk} = \|i(t)_j - i(t)_k\|_2$$

In this way, the relationship between active and non-active current is measured. These distances can also be interpreted as the entries of the distance similarity matrix $D \in \mathbb{R}^{w,w}$ for the points $i(t)_1, \ldots, i(t)_w$:

$$D = \begin{bmatrix} 0 & d_{12} & \cdots & d_{1w} \\ d_{21} & 0 & \cdots & d_{2w} \\ \vdots & \vdots & \ddots & \vdots \\ d_{w1} & d_{w2} & \cdots & 0 \end{bmatrix}$$

Finally, this matrix is used to calculate the Weighted Recurrence Graph matrix $WRG = [r_{jk}]$ with:

$$r_{jk} = \begin{cases} \delta & \text{if } \tau > \delta \\ \tau & \text{otherwise} \end{cases}$$

where $\tau = \left\lfloor \frac{d_{jk}}{\epsilon} \right\rfloor$ and the parameters δ and ϵ are chosen, in our case, both equal to 10.

This matrix is then used as input for the classifier, which in our case is a Convolutional Neural Network. The described algorithm is summarized in Figure 1.

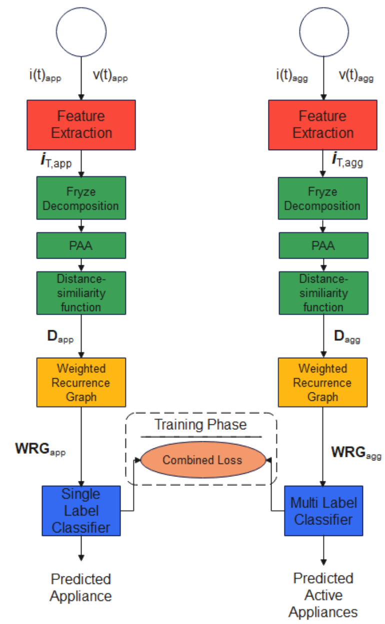

Figure 1: Block diagram of the proposed algorithm. The dotted block shows the training phase of the two classifiers, in which they learn by using the same combined loss function.

2.3 Combined Model

The main part of this paper is to give an idea of how the two problems can be approached from a multi-task perspective using a single algorithm. The idea was, therefore, to train two different classifiers, one for the single-label case and one for the multi-label case, which would share the information of the respective features in the learning phase. We propose two CNNs with the same structure for both problems, the only difference being the output size. Each classifier consist of four-stage CNN layer each with 16, 32, 64 and 128 feature maps, 1×1 strides and each layer uses a 3×3 filter size. The four layers are followed by a batch normalization layer and a Relu activation function, and the last layer is followed by an adaptive average pooling layer with an output size 1×1. The output layer consists of two fully connected layers with hidden size, 1024 and M, in the *Single Label* case, or $2M$, in the *Multi Label* case, where M is the number of available appliances. For both cases, the final predictions are obtained by applying the logarithmic softmax.

In the first case, the algorithm's output can be interpreted as a vector whose components represent each appliance's logarithm of the probability of activation or deactivation. Similarly, in the second case, the output represents two vectors containing the logarithm of the probability of each device being switched on or off.

To allow the parameters of the two models to share information during the learning phase, backpropagation was used to optimise a combined objective function defined by the equation:

$$\mathcal{L}_{comb}(\mathbf{m}, \hat{\mathbf{m}}, \mathbf{s}, \hat{\mathbf{s}}) = \mathcal{L}_{sl}(\mathbf{m}, \hat{\mathbf{m}}) + \mathcal{L}_{ml}(\mathbf{s}, \hat{\mathbf{s}}) \tag{3}$$

where $\mathcal{L}_{sl}$, $\mathcal{L}_{ml}$ represent the loss function in the *Single-Label* and in the *Multi-Label* case respectively. For both cases, the Negative Log-Likelihood Loss function is used [9].

The mini batch Stochastic Gradient Descent with a momentum of 0.9 and lerning rate of 0.001 is used to train the single-label model and an Adam optimizer with the same learning rate and betas of $(0.9, 0.98)$ is used to train the multi-label model. The combined model is then trained for 300 iterations using a batch size equal to 16 and to avoid the over-fitting we used an early-stopping with patience where the training phase is stopped once the validation performances don't change after 100 iterations.

The use of a combined loss to allow simultaneous training between the two models can be improved by introducing some penalty parameters to be added during the learning phase. Although the results of our approach, described in the Section 4, are interesting, the purpose of this paper is to propose a new approach to the problem and offer an idea for an improvable solution.

3 EVALUATION

3.1 Dataset

The proposed algorithm was tested on the PLAID dataset. This public dataset contains records, collected at high frequency (30kHz), measuring the individual current and voltage of 17 devices and the aggregate current and voltage of the combination of 13 of these. The data was collected at 65 locations in Pittsburgh, Pennsylvania (USA). In order to test our algorithm, 1046 activation and deactivation current and voltage measurements were extracted from 12 different individual appliances. This set of 2092 samples was used for the single-label part of our model. Each of these samples corresponds to the aggregated current and voltage measurements; when one or more devices are switched on or off consecutively, resulting in 2092 aggregated current and voltage samples used for the multi-label part of the model. The Figure 2 shows an example of the signals used, namely, a *Fan* current activation cycle, a *Vacuum* current activation cycle and the two appliances aggregated current cycle are illustrated.

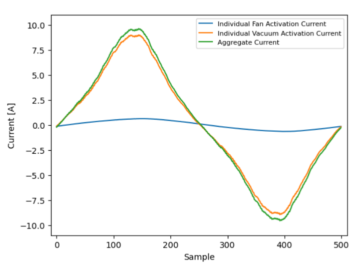

Figure 2: Extracted activation currents for two different events and the extracted aggregate current for these two events.

The Figure 3 summarises the distribution of extracted appliances and the number of active appliances in the 2092 samples.

3.2 Performance metrics

In order to evaluate the performance of our model, several metrics were used. The four main metrics used are Recall, Precision, and F1 score [11, 12]. These metrics were calculated to monitor the performance of the individual appliances in the two models to understand where one model could perform better than the other and study a way to combine the two strategies

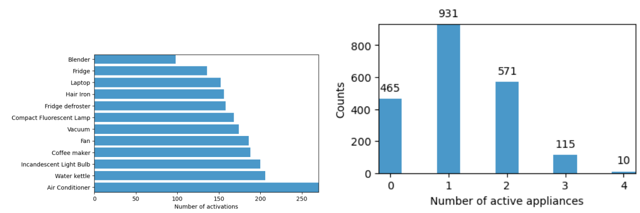

Figure 3: Appliances distribution on the extracted 2092 events (on the left) and the active appliances distribution (on the right).

$$\text{Recall}_{\text{macro}} = \frac{1}{M}\sum_{i=1}^{M}\text{Recall}^{(i)}$$

$$\text{Precision}_{\text{macro}} = \frac{1}{M}\sum_{i=1}^{M}\text{Precision}^{(i)} \qquad \text{F1}_{\text{macro}} = \frac{1}{M}\sum_{i=1}^{M}\text{F1}^{(i)}$$

where $\text{Recall}^{(i)}$, $\text{Precision}^{(i)}$, $\text{F1}^{(i)}$ are the metrics computed for the i-th appliance and M is the number of total appliances.

3.3 Training and Testing

Our approach is benchmarked using the multi-label stratified 3-fold Cross Validation; in this way, in addition, to avoid overfitting, the label percentages in each fold are fixed.

Although the classifiers' training phase takes place simultaneously and optimises the same loss function, it should be emphasised that the single-label classifier is trained using only the features extracted from the current and voltage signals of the appliance switched on or off. In contrast, the multi-label classifier uses the features from the aggregated signals.

During the test phase, predictions are computed considering the maximum logarithmic softmax output, and 30% of the data for each fold is used. The predictions of each fold are then merged into a single sequence to assess average performances.

4 RESULTS

The Figure 4 represents the performances for each appliance of the two simultaneously trained models, and it contains the metrics described in the previous Section.

As can be seen by looking at the values of the metrics, for some appliances, there is a difference in performance between the two tasks (for instance, *Fan* or *Blender*); our idea is that by properly combining the two classification strategies the model with the better performance can help increase the performance of the other.

Although not comparable, the two classifiers have similar good characteristics from the point of view of general performance. The Table 1 summarises the general performance.

A miss-classification error analysis was also carried out to understand whether the switching on or off of some appliance could cause errors not common to both models and, in this way, to think about how the model that makes mistakes could learn from the model that does not make mistakes.

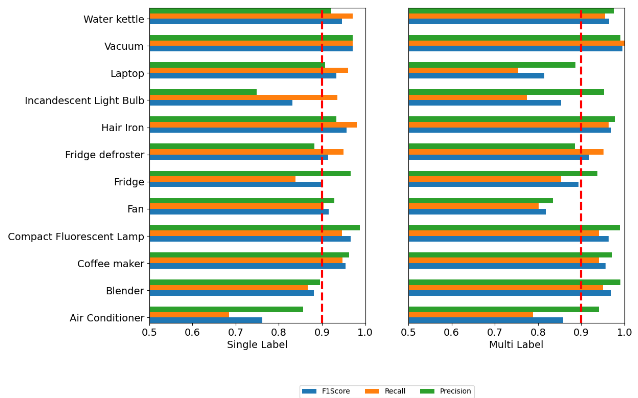

Figure 4: Per-appliance F1 score, Recall and Precision on the PLAID dataset. *Single Label* (left), *Multi Label* (right).

Table 1: Generalisation performances between *Single Label* model and *Multi Label* model on the PLAID dataset.

	Single Label Model	Multi Label Model
Recall$_{\mathbf{macro}}$	0.913	0.889
Precision$_{\mathbf{macro}}$	0.913	0.944
F1$_{\mathbf{macro}}$	0.910	0.915

Figure 5 illustrates the number of these errors, they are divided into errors committed only by one model and common errors; more precisely, errors committed only by the single-label model but not by the multi-label model (*Only SingleLabel Errors*), those committed by the multi-label model but not by the single-label model (*Only MultiLabel Errors*), and those committed by both algorithms (*Common Errors*) are counted.

It can be seen from the graph that in the case of the *Laptop*, for example, the multi-label model makes many more errors than the single-label model; appropriate information sharing between the two models could lead to a reduction in the number of errors.

The confusion matrix related to the predictions of the single-label model, illustrated in Figure 6, points out, for example, that a good number of errors made by the classifier are related to the *Air Conditioner* and also that most of these are related to deactivation events. In this case, the multi-label model could help the single-label classification.

5 CONCLUSION

In this paper, we proposed an approach to address two types of appliance recognition problems simultaneously; single and multi-label appliance recognition. The paper aims to motivate the search for a model capable of sharing information between two classifiers under training to improve appliance recognition.

Our approach, albeit in a primitive form, was tested on the public dataset PLAID to prove its effectiveness. In fact, from the results obtained, it seems that the proposed method can offer an excellent basis to address and solve the problem. In addition, we think the proposed approach could also be used to improve the event detection algorithm, which will likely pave the way toward end-to-end NILM.

The work presented in this paper is intended to encourage the combined approach of two interrelated problems; therefore, it is clear that this work can be extended and improved. One of the strategies to be applied for future work is to consider a linear combination of the two loss functions instead of the sum to train the classifiers. The coefficients of the combination could be considered as weights or penalty factors or not less than trainable parameters. Such approaches, moreover, could be validated using different datasets in addition to the one proposed, e.g. a dataset that allows the use of Time-Series cross-validation and enables an approach much closer to the real case. Finally, a post-processing step could be helpful to improve the performance of the proposed model by considering the predictions of one of the two classifiers and checking that they are somehow compatible with the predictions of the other classifier.

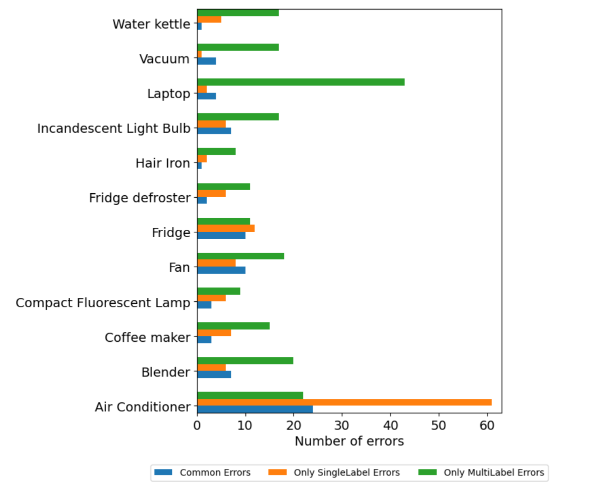

Figure 5: Number of incorrect predictions (during the testing phase) divided into errors committed by the *Single Label* model only, the *Multi Label* model only and those committed by both models (*Common Errors*).

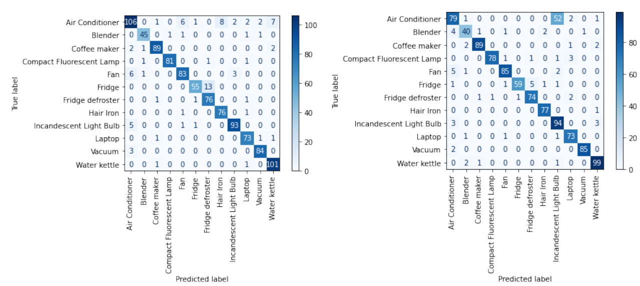

Figure 6: Confusion matrices of *Single Label* model predictions. On the left only the predictions for activation events are considered, on the right only the predictions for deactivation events.

ACKNOWLEDGMENTS

Lucas Pereira received funding from FCT under grants CEECIND/01179/2017, and UIDB/50009/2020.

REFERENCES

[1] Deyvison de Paiva Penha and Adriana Rosa Garcez Castro. 2018. Home appliance identification for NILM systems based on deep neural networks. *Int. J. Artif. Intell. Appl* 9 (2018), 69–80.

[2] Anthony Faustine and Lucas Pereira. 2020. Improved appliance classification in non-intrusive load monitoring using weighted recurrence graph and convolutional neural networks. *Energies* 13, 13 (2020), 3374.

[3] Anthony Faustine and Lucas Pereira. 2020. Multi-label learning for appliance recognition in NILM using Fryze-current decomposition and convolutional neural network. *Energies* 13, 16 (2020), 4154.

[4] Anthony Faustine, Lucas Pereira, and Christoph Klemenjak. 2020. Adaptive Weighted Recurrence Graphs for Appliance Recognition in Non-Intrusive Load Monitoring. *IEEE Transactions on Smart Grid* (2020), 1–1. https://doi.org/10.1109/TSG.2020.3010621

[5] Jack Kelly and William Knottenbelt. 2015. Neural nilm: Deep neural networks applied to energy disaggregation. In *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments*. 55–64.

[6] Franco Marchesoni-Acland, Camilo Mariño, Elías Masquil, Pablo Masaferro, and Alicia Fernández. 2020. End-to-end NILM System Using High Frequency Data and Neural Networks. https://doi.org/10.48550/ARXIV.2004.13905

[7] Luca Massidda, Marino Marrocu, and Simone Manca. 2020. Non-intrusive load disaggregation by convolutional neural network and multilabel classification. *Applied Sciences* 10, 4 (2020), 1454.

[8] Roberto Medico, Leen De Baets, Jingkun Gao, Suman Giri, Emre Kara, Tom Dhaene, Chris Develder, Mario Bergés, and Dirk Deschrijver. 2020. A Voltage and Current Measurement Dataset for Plug Load Appliance Identification in Households. *Scientific Data* 7, 1 (Feb. 2020), 1–10. https://doi.org/10.1038/s41597-020-0389-7

[9] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre GR Day, Clint Richardson, Charles K Fisher, and David J Schwab. 2019. A high-bias, low-variance introduction to machine learning for physicists. *Physics reports* 810 (2019), 32.

[10] Christoforos Nalmpantis and Dimitris Vrakas. 2020. On time series representations for multi-label NILM. *Neural Computing and Applications* 32, 23 (2020), 17275–17290.

[11] Lucas Pereira and Nuno Nunes. 2017. A Comparison of Performance Metrics for Event Classification in Non-Intrusive Load Monitoring. In *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 159–164. https://doi.org/10.1109/SmartGridComm.2017.8340682

[12] Lucas Pereira and Nuno Nunes. 2020. An Empirical Exploration of Performance Metrics for Event Detection Algorithms in Non-Intrusive Load Monitoring. *Sustainable Cities and Society* 62 (Nov. 2020), 102399. https://doi.org/10.1016/j.scs.2020.102399

[13] Muhammad Asif Ali Rehmani, Saad Aslam, Shafiqur Rahman Tito, Snjezana Soltic, Pieter Nieuwoudt, Neel Pandey, and Mollah Daud Ahmed. 2021. Power Profile and Thresholding Assisted Multi-Label NILM Classification. *Energies* 14, 22 (2021), 7609.

[14] Dawit Fekadu Teshome, T. D. Huang, and Kuo-Lung Lian. 2016. Distinctive Load Feature Extraction Based on Fryze's Time-Domain Power Theory. *IEEE Power and Energy Technology Systems Journal* 3, 2 (2016), 60–70. https://doi.org/10.1109/JPETS.2016.2559507

[15] Xin Wu, Dian Jiao, Kaixin Liang, and Xiao Han. 2019. A fast online load identification algorithm based on VI characteristics of high-frequency data under user operational constraints. *Energy* 188 (2019), 116012.