A Multiple Instance Regression Approach to Electrical Load Disaggregation

Luca Serafini, Giulia Tanoni, Emanuele Principi, Susanna Spinsante, and Stefano Squartini
Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy
Email: {l.serafini,e.principi,s.spinsante,s.squartini}@staff.univpm.it, g.tanoni@pm.univpm.it

***Abstract*—Non-Intrusive Load Monitoring (NILM) provides detailed information on the consumption of individual appliances in a building and represents an effective method to reduce the electricity consumed in the residential sector. Supervised Deep Learning approaches have achieved the state-of-the-art for NILM but require knowledge of *strongly* labeled data, i.e., annotated at the sample level. This data is costly to obtain since it requires multiple sensors to measure electrical quantities and the involvement of the end-users.**

This work proposes a Multiple Instance Regression approach to NILM using a Convolutional Recurrent Neural Network (CRNN) to reduce the amount of strongly labeled data required for training and improve performance. *Instances* of strongly labeled data are here represented by raw samples of active power and are aggregated into *bags* containing weak information represented by the average power consumption in a bag. Using this information, the network is trained to disaggregate appliances' power profiles with sample resolution. The results obtained on the UK-DALE dataset demonstrated the approach's effectiveness in reducing the labeling cost and improving the performance: the average Mean Absolute Error reduces by 3.06 W when weak information is used in the CRNN and by 8.88 W compared to the Sequence-to-Point method.

***Index Terms*—Non-Intrusive Load Monitoring, Electrical Load Disaggregation, Weak Supervision, Deep Learning.**

I. Introduction

Non-Intrusive Load Monitoring (NILM) [1] was proposed by Hart in 1980 to extract appliance consumption profiles from the knowledge of the aggregate measurements. This way, instead of individually monitoring each appliance inside a building, the number of sensors for measuring their active power is reduced to one. NILM is very effective in supporting energy-saving since it provides detailed knowledge of load power consumption to end-users. In [2], it has been reported that real-time and appliance-level consumption feedback leads to a reduction of energy consumption up to 12%.

NILM is based on advanced algorithms for extracting detailed consumption information. During the last decades, Signal Processing [3], [4], Single-Channel Source Separation [5], and Machine Learning (ML) [6]–[12] techniques have been largely employed for this task. ML-based approaches, particularly Deep Neural Networks, have proven to be the most effective and have reached the state-of-the-art. For load disaggregation, i.e., direct estimation of the appliance active power, these approaches are based on fully supervised learning [7]–[13]. Kelly et al. [7] firstly proposed three different architectures to estimate the appliance power consumption from sequences of aggregate samples. In [8], the sequence-to-point approach, based on a Convolutional Neural Network (CNN), has been proposed by training the network to estimate the middle point of the target window. Recently, in [9], a Variational Auto-Encoder has been presented to improve multi-state appliance disaggregation generalization capability. Piccialli et al. [13] proposed a double-branched architecture, one for regression, with an attention mechanism, and one for classification. The network is trained by minimizing the sum of the two losses on the target sequence. Also, Laouali et al. [10] proposed a double-branched approach exploiting the active and reactive aggregate power and the ApproxHull data selection strategy. Recently, in [11], a CNN and a Long Short Term Memory network have been proposed with the attention mechanism, improving the disaggregation for complex-state devices. Nalmpantis et al. [12] proposed a Neural Fourier Energy Disaggregation approach to reduce the complexity of the network while keeping the performance unchanged.

The described methods are based on fully or *strongly* supervised learning strategies that require the knowledge of a large quantity of aggregate and appliance-level measurements annotated sample by sample. Labeling is a monetary and human costly procedure in NILM due to the electrical sensors required to record individual appliance consumption and the end-user involvement during the acquisition phase. Unlike strongly supervised methods, *weak* supervision refers to the learning strategies that do not need complete knowledge of data [14]. For example, Semi-Supervised Learning consists in learning with incomplete supervision since it employs unlabeled data. In contrast, Multiple Instance Learning [15] and Multiple Instance Regression (MIR) [16] consist in learning with *inexact* supervision since they use coarse-grained information as ground-truth data. Up to the authors' knowledge, weak supervision has been proposed in the classification task to estimate the states of the appliances exploiting unlabeled data [17], [18] while it has never been employed to estimate the appliance's active power.

In this work, we propose a MIR approach to electrical load disaggregation. Generally, in MIR, the objective is learning to predict multiple real-valued variables by using weak labels

This work was partially funded by the Department of Information Engineering of Università Politecnica delle Marche, under the project "Un approccio integrato per soluzioni innovative ed eco-sostenibili di trasporto merci nella logistica emergenziale e dell'ultimo miglio" (RSA-B 2020) and by the Marche Region in implementation of the financial programme POR MARCHE FESR 2014-2020, project "MIRACLE" (Marche Innovation and Research fAcilities for Connected and sustainable Living Environments), CUP B28I19000330007.

for training. Here, real-valued variables are samples of the active power of an appliance and represent strong labels. In contrast, a weak label is the average value of a segment of appliance's active power values. Referring to the MIR terminology, samples here represent the *instances*, and segments represent *bags*. The proposed method employs both weak and strong labels to train a Convolutional Recurrent Neural Network (CRNN) to perform load disaggregation. NILM, thus, is modeled as a MIR problem, and we exploit the presence of weakly labeled data to reduce the quantity of strongly labeled data required in fully supervised approaches and improve the performance. The UK-DALE dataset [19] has been used to perform the experiments and prove the effectiveness of the method. The experiments have been conducted by varying the composition of the training dataset, each time changing the amount of strongly labeled data while keeping fixed the quantity of weakly labeled data. The proposed method has been compared to a CRNN trained only on strongly labeled data (CRNN-Strong) and to the Sequence-to-Point network (Seq2Point) presented in [8]. The obtained results demonstrate the increased generalization ability on unseen conditions and the reduction of disaggregation error provided by the proposed method compared to the fully supervised strategy. The most significant improvements have been obtained when the amount of strongly labeled data is lower than the amount of weakly labeled data. The obtained results allow concluding that the proposed method is able to achieve superior performance with respect to the comparative algorithms using a significantly lower amount of strongly labeled data and exploiting weak labels, i.e., coarser information that requires significantly less effort for annotation. Up to our knowledge, this is the first work in which electrical load disaggregation has been addressed by using weak supervision and MIR.

The outline of the paper is the following. Section II defines the load disaggregation problem and explains in detail the proposed method; Section III describes the experimental setup and presents the obtained results. Finally, Section IV concludes the paper and presents future works.

II. Proposed Method

The total active power consumption of a building is given by the sum of the individual active powers consumed by all the appliances inside it, and it can be expressed as follows:

$$y(t) = \sum_{n=1}^{N} x_n(t) + e(t), \tag{1}$$

where $x_n(t)$ represents the active power of appliance n at the instant t, N is the total number of the appliances, and $e(t)$ indicates the measurement noise and the contribution of the appliances whose consumption is not of interest.

Load disaggregation aims at extracting the active power consumption profile of a specific appliance n, given only the power reading of the mains $y(t)$. In this framework, load disaggregation is formulated as a denoising task, as in several previous works on the topic [7], [20]. Considering the active power of appliance $\bar{n}$, equation (1) can be expressed alternatively as:

$$y(t) = x_{\bar{n}}(t) + v_{\bar{n}}(t), \quad v_{\bar{n}}(t) = \sum_{\substack{n=1 \\ n \neq \bar{n}}}^{N} x_n(t) + e(t). \tag{2}$$

where $v_{\bar{n}}(t)$ represents the total noise term for appliance $\bar{n}$. In this formulation, load disaggregation consists in removing the noise contribution $v_{\bar{n}}(t)$ from $y(t)$ for each appliance.

The proposed method addresses load disaggregation as a MIR problem and uses a CRNN for denoising. Based on MIR, the individual samples of the signals are *instances* that are grouped into *bags*. Each bag is represented by a real-valued label based on the values of the instances contained in it. Since the information carried by bag target values is coarse, they are also referred to as *weak* while, in contrast, instance labels are also referred to as *strong* because of their higher information resolution.

In the proposed framework, instances are represented by samples of power reading of the mains $y(t)$. To introduce the concept of a *bag of instances*, we divide $y(t)$ into windows of fixed length L and overlapped by $P < L$ samples, and we define a bag j as the j-th window of $y(t)$ as follows:

$$\mathbf{y}_j = [y(j(L-P)), \ldots, y(j(L-P)+L-1)]^T. \tag{3}$$

Omitting n for simplicity of notation, the strong labels for bag $\mathbf{y}_j$ of a generic appliance are represented by the ground-truth data $\mathbf{x}_j = [x(j(L-P)), \ldots, x(j(L-P)+L-1)]^T$.

The weak label of a bag depends on the strong labels of the instances within the bag itself. Generally, the weak label is obtained from the instance labels by using a pooling function. Several alternatives have been proposed in the literature for classification and regression [15]. In this work, the weak label w_j related to bag $\mathbf{y}_j$ and a generic appliance is a scalar quantity calculated as the arithmetic average of the instance labels:

$$w_j = \frac{1}{L} \sum_{l=0}^{L-1} x_n(j(L-P)+l). \tag{4}$$

With the above definitions, it is possible to formulate load disaggregation using MIR more formally. Denoting with

$$\begin{aligned} \mathcal{T} = \{&(\mathbf{y}_1, w_1, \mathbf{x}_1), \ldots, (\mathbf{y}_M, w_M, \mathbf{x}_M), \\ &(\mathbf{y}_{M+1}, w_{M+1}), \ldots, (\mathbf{y}_{M+K}, w_{M+K})\}, \end{aligned} \tag{5}$$

a set of $M+K$ bags, in which M are annotated with strong and weak labels and K only with weak labels, the goal is to learn a mapping function $\mathbf{f} : \mathbb{R}^L \to \mathbb{R}^L$ from $\mathcal{T}$ for estimating the active power $\mathbf{x}$ of an appliance given a bag of unknown aggregate power $\mathbf{y}$. The mapping function $\mathbf{f}(\cdot)$ is represented by a CRNN, described in the following section.

A. Neural Network Architecture

As previously stated, we address load disaggregation by using a CRNN [21]. Specifically, we train a different CRNN for each appliance of interest. The network takes a bag $\mathbf{y}$ of

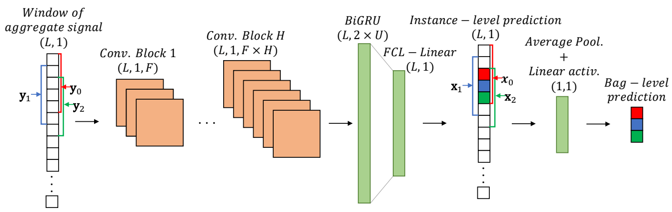


Fig. 1. The proposed architecture.

aggregate power as input and has two outputs: a strong-level output and a weak-level output. The first provides an estimate of the active power $\hat{\mathbf{x}}$. Supposing that L is odd, and P is even, since the aggregate signal is processed in partially overlapped windows and $\mathbf{y}$ and $\hat{\mathbf{x}}$ are of the same length L, we retain only the $L-P$ central values of the output. The entire output sequence is reconstructed by joining the individual output segments. The weak-level output is represented by the average of the instance-level predictions, consistently with equation (4).

The network contains H convolutional blocks, where each of them includes: a 2D convolutional layer with F filters and kernel size K, a layer for batch normalization, an activation layer that exploits the Rectified Linear Unit activation function, a max-pooling layer, and a dropout layer with rate p. The output of the convolutional blocks is connected to the recurrent part of the CRNN, which is represented by a Bidirectional layer of U Gated Recurrent Units (GRU). The last layer is a Fully-Connected Layer with a linear activation function for generating the strong-level predictions associated with the input window. The strong-level output is further processed by an average pooling layer, followed by a linear activation function in order to generate the weak-level prediction. The CRNN architecture proposed in this paper is depicted in Fig. 1.

B. Learning

Given a set of annotated bags $\mathcal{T}$, the network is trained by using a loss $\mathcal{L} = \mathcal{L}_s + \lambda\mathcal{L}_w$ given by the weighted sum of the loss associated with the strong-level output $\mathcal{L}_s$, and the one related to the weak-level output $\mathcal{L}_w$. The term λ is a weight that balances the contribution of the two losses. Both $\mathcal{L}_s$ and $\mathcal{L}_w$ are calculated as the Mean Squared Error between the related prediction and the target. Considering a mini-batch containing J bags and a generic appliance, the two losses are calculated as follows:

$$\mathcal{L}_s = \frac{1}{J \cdot L} \sum_{j=0}^{J-1} \sum_{l=0}^{L-1} \left[x(j(L-P)+l) - \hat{x}(j(L-P)+l)\right]^2, \tag{6}$$

$$\mathcal{L}_w = \frac{1}{J} \sum_{j=0}^{J-1} \left(w_j - \hat{w}_j\right)^2. \tag{7}$$

III. Experiments

This section describes the experiments conducted to evaluate the proposed method.

A. Datasets

The experiments have been carried out by using the UK-DALE dataset [19]. It is a publicly available dataset that contains data related to five different houses in the UK, where aggregate power is sampled at 1 Hz, while appliance-level measurements at 1/6 Hz. The appliances considered in the experiments are Microwave (MW), Fridge (FR), Dishwasher (DW), Washing Machine (WM), and Kettle (KE). All houses were included but, for houses 3 and 4, only Kettle and Fridge were considered. The periods considered for each house are 2013/04/12-2015/01/05 for house 1, 2013/05/22-2013/10/10 for house 2, 2013/02/27-2013/04/08 for house 3, 2013/03/09-2013/10/01 for house 4, and 2014/06/29-2014/11/13 for house 5.

B. Evaluation metrics

The metrics used to evaluate the performance of the method are the Mean Absolute Error (MAE) and the Normalized Error in assigned Power (NEP) [22]. Both metrics are calculated for each appliance individually. MAE and NEP are defined as:

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} \left|x(t) - \hat{x}(t)\right|, \quad \text{NEP} = T \cdot \frac{\text{MAE}}{\sum_{t=1}^{T} x(t)}, \tag{8}$$

where $\hat{x}(t)$ is the power predicted by the network, $x(t)$ is the corresponding ground-truth value, and T is the number of samples of the segment under evaluation. Basically, NEP is the MAE normalized by the total appliance's energy consumption, and it allows to evaluate the importance of the error based on the appliance operating characteristics.

C. Experimental procedure

We downsampled the aggregate active power to 6 s and we aligned it to the appliance readings using NILMTK [22]. Weak ground-truth labels have been created by using equation (4). The experiments have been conducted on an *unseen* scenario, using house 2 data only for testing and data of the other houses to train and validate the model. For each appliance, we first divided the original dataset into training (1,200,000 samples), validation (150,000 samples), and testing sets (2,100,000 for Kettle and 1,700,000 samples for the other appliances).

Evaluation has been conducted in multiple training conditions, each characterized by a different amount of strong labels. We considered two extreme situations, one where the amount of strong labels is very scarce, i.e., 5% the number of strong labels in the training set, and one where it is large, i.e., 100% the amount in the training set. Moreover, we considered three intermediate values, 20%, 40%, and 80%, thus each time doubling the amount of strong labels for training. The number of weak labels, on the other hand, is always the same. For each appliance and training condition, we standardized the aggregate signal by using mean and standard deviation calculated from the training set and applied min-max normalization to the target values.

For each training condition, we also evaluated the performance of the Sequence-to-Point network proposed in [8], and of the same CRNN network depicted in Fig. 1 but without the bag-level output. Thus, training has been performed only on strong labels in these cases.

Both for the proposed and the comparative methods, we trained a different network for each appliance of interest by setting the maximum number of *epochs* to 1000 and using the Early-Stopping regularization technique with *patience* equal to 15 epochs. During the learning process, we used the Adaptive Moment Estimation optimizer [23], with a learning rate equal to 0.001, β_1 and β_2 equal to 0.9 and 0.999, respectively, and ϵ equal to 10^{-7}. The loss weight λ has been set to 1. Training has been performed in mini-batches, where the batch size has been determined on the validation set. The length L of the window is different for each appliance, and it has been determined by evaluating the performance on the validation set of the values reported in [8], [20]. The determined values are 289 for Microwave, 1025 for Washing Machine, and 599 for Fridge, Dishwasher, and Kettle. Differently, the amount of overlap P is equal for all the appliances and has been set to $(L-1)$. This means that for each input bag, only the central value of the output is retained.

All the methods have been implemented in Python using Tensorflow and Keras [24].

D. Results

The results obtained for each appliance and training condition are reported in Table II. The proposed method is indicated with "Proposed", the CRNN trained only with strong labels with "CRNN-Strong", while the Sequence-to-Point network with "Seq2Point".

The obtained results show that regardless of the percentage of strongly labeled data used for training, the proposed method based on weak labels is able to outperform the comparative algorithms. Compared to CRNN-Strong, MAE reduces by 3.06 W on average, while compared to Seq2Point by 8.88 W. Similarly, NEP reduces by 9.97 percentage points (pp) compared to CRNN-Strong and by 30.59 pp compared to Seq2Point.

Observing the performance for the different percentages of strong labels, the most remarkable improvement occurs when the percentage of strongly labeled data is low, i.e., 5%, 20%, and 40%, both when the proposed method is compared to CRNN-Strong and when it is compared to Seq2Point. This behavior confirms that weak labels contribute the most to improving the performance when the amount of strongly labeled data is scarce compared to weakly labeled data. Another remarkable advantage of the proposed method is the reduction of strongly labeled data quantity required to obtain the lowest error. As it can be seen for Microwave, Fridge and Dishwasher, the lowest error among all the percentages is achieved with weakly labeled data and when the quantity of strongly labeled data is only 20% while for Washing Machine only the 5%.

TABLE I

CRNN HYPERPARAMETERS.

Hyperparameter	Symbol	Value
Number of convolutional blocks	H	3
Number of filters	F	[32, 64, 128]
Kernel size	K	5
Number of GRU units	U	64
Dropout rate	p	0.1

TABLE II

RESULTS OBTAINED FOR THE DIFFERENT TRAINING CONDITIONS AND ADDRESSED METHODS. BEST RESULTS FOR EACH APPLIANCE AND PERCENTAGE OF STRONG LABELS ARE REPORTED IN BOLD.

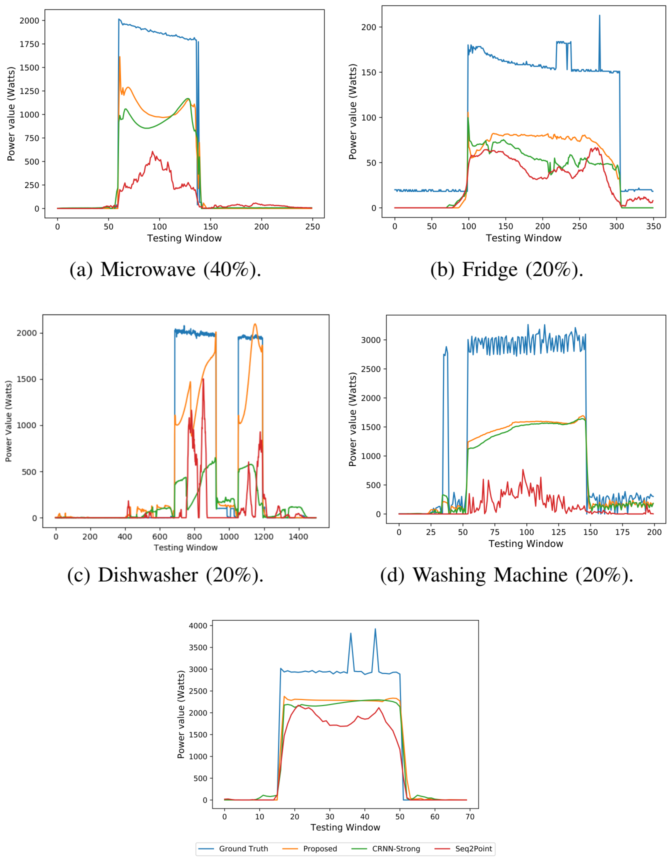
% Strong	Model	Metric	Appliance: MW	FR	DW	WM	KE	Average
5%	Proposed	MAE (W)	**11.17**	**49.89**	**20.55**	**10.93**	**12.57**	**21.02**
		NEP (%)	**113.82**	**59.69**	**48.98**	**70.56**	**42.52**	**67.11**
	CRNN-Strong	MAE (W)	11.38	50.04	32.13	12.05	13.46	23.81
		NEP (%)	115.95	59.86	76.57	77.78	45.53	75.14
	Seq2Point	MAE (W)	15.31	66.07	48.96	14.45	28.72	34.70
		NEP (%)	155.92	79.05	116.69	93.23	97.14	108.41
20%	Proposed	MAE (W)	**8.16**	**46.93**	**20.08**	**11.44**	**12.37**	**19.80**
		NEP (%)	**83.10**	**56.15**	**47.86**	**73.82**	**41.84**	**60.55**
	CRNN-Strong	MAE (W)	8.32	47.16	37.28	12.65	14.96	24.07
		NEP (%)	84.75	56.42	88.85	81.65	50.59	72.45
	Seq2Point	MAE (W)	11.15	66.15	31.57	12.99	29.01	30.17
		NEP (%)	113.54	79.14	75.24	83.83	98.12	89.97
40%	Proposed	MAE (W)	**10.61**	**48.07**	**28.12**	**11.70**	**11.52**	**22.00**
		NEP (%)	**108.09**	**57.52**	**67.02**	**75.48**	**38.97**	**69.42**
	CRNN-Strong	MAE (W)	15.29	53.79	33.33	12.49	12.63	25.51
		NEP (%)	155.83	64.35	79.44	80.63	42.71	84.59
	Seq2Point	MAE (W)	13.09	65.38	46.42	23.00	14.71	32.52
		NEP (%)	133.37	78.22	110.63	148.43	49.74	104.08
80%	Proposed	MAE (W)	9.90	**50.95**	**26.62**	**12.19**	**10.54**	**22.04**
		NEP (%)	101.00	**60.96**	**63.45**	**78.64**	**35.66**	**67.94**
	CRNN-Strong	MAE (W)	**9.55**	52.03	33.09	12.53	15.33	24.51
		NEP (%)	**97.00**	62.25	78.86	80.88	51.86	74.17
	Seq2Point	MAE (W)	17.11	53.98	29.60	14.61	13.17	25.69
		NEP (%)	174.23	64.58	70.55	94.28	44.53	89.63
100%	Proposed	MAE (W)	**10.89**	49.97	**30.40**	13.02	**11.82**	**23.22**
		NEP (%)	**110.96**	59.79	**72.46**	84.02	**39.97**	**73.44**
	CRNN-Strong	MAE (W)	12.70	**49.90**	37.14	**12.72**	14.85	25.46
		NEP (%)	129.34	**59.70**	88.52	**82.12**	50.24	81.98
	Seq2Point	MAE (W)	18.03	65.85	32.18	17.31	13.55	29.38
		NEP (%)	183.67	78.79	76.71	111.68	45.84	99.34
Average	Proposed	MAE (W)	**10.15**	**49.16**	**25.15**	**11.86**	**11.76**	**21.62**
		NEP (%)	**103.39**	**58.82**	**59.95**	**76.50**	**39.79**	**67.69**
	CRNN-Strong	MAE (W)	11.45	50.58	34.59	12.49	14.25	24.67
		NEP (%)	116.57	60.52	82.45	80.61	48.19	77.67
	Seq2Point	MAE (W)	14.94	63.49	37.75	16.47	19.83	30.49
		NEP (%)	152.15	75.96	89.96	106.29	67.07	98.29

A closer look at the behavior for the different appliances shows that in the majority of the cases, the performance of the proposed method is superior to the comparative methods. The only exceptions are Microwave when the percentage of strong labels is 80%, and Fridge and Washing Machine when the percentage is 100%. Note, however, that the MAE difference is below 0.5 W and that this occurs when the amount of strongly and weakly labeled data is comparable: in this case, the influence of weak labels is less, a behavior that could have been expected.

Fig. 2 shows the ground-truth and the estimated active power for the proposed and comparative methods when training is performed with different percentages of strongly labeled data. The plots confirm the obtained results, as the active power outputs produced using the proposed method are closer to the ground-truth.

IV. CONCLUSION AND FUTURE WORKS

This paper presented a load disaggregation method based on Multiple Instance Regression and a Deep Neural Network represented by a Convolutional Recurrent Neural Network. The learning strategy is able to exploit data annotated with both strong and weak labels thus, it is able to use coarser annotations that are intrinsically less expensive to obtain.



(a) Microwave (40%). (b) Fridge (20%). (c) Dishwasher (20%). (d) Washing Machine (20%). (e) Kettle (20%).

Fig. 2. Ground-truth and estimated active power for the proposed and comparative methods for different percentages of strongly labeled data (shown in brackets).

In the experiments, we evaluated whether the information carried out by the weak labels is able to provide significant performance improvements for load disaggregation. The UK-DALE dataset has been used in the experiments, and different training conditions with various percentages of strongly labeled data and fixed amount of weakly-annotated data have been considered. The proposed method has been evaluated in unseen testing conditions, and it has been compared to two methods: a CRNN architecture missing the capability to exploit weak labels, and the Sequence-to-Point method presented in [8]. The experiments showed that the proposed method provides significant performance improvements, as the obtained results showed that on average it outperforms the comparative algorithms. This demonstrated the possibility to achieve superior performance by using data annotated with weak labels, that are intrinsically less costly to obtain.

Future developments will address also the possibility to use weak labels together with transfer learning techniques and explore different neural network architectures.

References

[1] G. Hart, "Nonintrusive appliance load monitoring," *Proc. of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

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

[3] B. Zhao, K. He, L. Stankovic, and V. Stankovic, "Improving event-based non-intrusive load monitoring using graph signal processing," *IEEE Access*, vol. 6, pp. 53 944–53 959, 2018.

[4] A. Moradzadeh, O. Sadeghian, K. Pourhossein, B. Mohammadi-Ivatloo, and A. Anvari-Moghaddam, "Improving residential load disaggregation for sustainable development of energy via principal component analysis," *Sustainability*, vol. 12, no. 8, 2020.

[5] A. Rahimpour, H. Qi, D. Fugate, and T. Kuruganti, "Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint," *IEEE Trans. Power Syst.*, vol. 32, no. 6, p. 4430–4441, 2017.

[6] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, "Non-intrusive load monitoring by using active and reactive power in additive factorial hidden markov models," *Applied Energy*, vol. 208, pp. 1590–1607, 2017.

[7] J. Kelly and W. Knottenbelt, "Neural NILM: Deep Neural Networks Applied to Energy Disaggregation," in *Proc. of BuildSys*, New York, USA, Nov. 4–5 2015, pp. 55–64.

[8] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, "Sequence-to-point learning with neural networks for non-intrusive load monitoring," in *Proc. of the 32nd AAAI Conf. on Artificial Intelligence*, New Orleans, LA, USA, Feb. 2-7 2018, pp. 2604–2611.

[9] A. Langevin, M.-A. Carbonneau, M. Cheriet, and G. Gagnon, "Energy disaggregation using variational autoencoders," *Energy and Buildings*, vol. 254, p. 111623, 2022.

[10] I. Laouali, A. Ruano, M. d. G. Ruano, S. D. Bennani, and H. E. Fadili, "Non-Intrusive Load Monitoring of Household Devices Using a Hybrid Deep Learning Model through Convex Hull-Based Data Selection," *Energies*, vol. 15, no. 3, 2022.

[11] Q. Liu *et al.*, "Improving wireless indoor non-intrusive load disaggregation using attention-based deep learning networks," *Physical Communication*, vol. 51, p. 101584, 2022.

[12] C. Nalmpantis, N. Virtsionis Gkalinikis, and D. Vrakas, "Neural Fourier Energy Disaggregation," *Sensors*, vol. 22, no. 2, 2022.

[13] V. Piccialli and A. M. Sudoso, "Improving Non-Intrusive Load Disaggregation through an Attention-Based Deep Neural Network," *Energies*, vol. 14, no. 4, p. 847, 2021.

[14] Z. H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 5, no. 1, pp. 44–53, 2018.

[15] J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, pp. 81–105, 2013.

[16] S. Ray and D. Page, "Multiple Instance Regression," in *Proc. of the 18th Int. Conf. on Machine Learning*, 2001, pp. 425–432.

[17] Y. Yang, J. Zhong, W. Li, T. A. Gulliver, and S. Li, "Semisupervised multilabel deep learning based nonintrusive load monitoring in smart grids," *IEEE Trans. Ind. Informat.*, vol. 16, no. 11, pp. 6892–6902, 2020.

[18] N. Miao, S. Zhao, Q. Shi, and R. Zhang, "Non-Intrusive Load Disaggregation Using Semi-Supervised Learning Method," in *Proc. of SPAC*, Guangzhou Shi, China, Dec. 20-23 2019, pp. 17–22.

[19] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Scientific Data*, vol. 2, no. 150007, 2015.

[20] R. Bonfigli, A. Felicetti, E. Principi, M. Fagiani, S. Squartini, and F. Piazza, "Denoising autoencoders for non-intrusive load monitoring: Improvements and comparative evaluation," *Energy and Buildings*, vol. 158, pp. 1461–1474, 2018.

[21] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. of EMNL*, Doha, Qatar, 2014, pp. 1724–1734.

[22] N. Batra *et al.*, "NILMTK: An open source toolkit for non-intrusive load monitoring," in *Proc. of the 5th Int. Conf. on Future Energy Systems*, Cambridge, UK, 2014, pp. 265–276.

[23] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Int. Conf. on Learning Representations*, 12 2014.

[24] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th {USENIX} Symp. Oper. Syst. Design Implement.*, 2016, pp. 265–283.