Multi-label Deep Convolutional Transform Learning for Non-intrusive Load Monitoring

Shikha Singh, Emilie Chouzenoux, Giovanni Chierchia, Angshul Majumdar

Multi-label Deep Convolutional Transform Learning for Non-intrusive Load Monitoring

SHIKHA SINGH, Indraprastha Institute of Information Technology, India
EMILIE CHOUZENOUX, Inria Saclay, OPIS, Center for Visual Computing, France
GIOVANNI CHIERCHIA, Université Gustave Eiffel, ESIEE Paris, France
ANGSHUL MAJUMDAR, Indraprastha Institute of Information Technology, India

The objective of this letter is to propose a novel computational method to learn the state of an appliance (ON / OFF) given the aggregate power consumption recorded by the smart-meter. We formulate a multi-label classification problem where the classes correspond to the appliances. The proposed approach is based on our recently introduced framework of convolutional transform learning. We propose a deep supervised version of it relying on an original multi-label cost. Comparisons with state-of-the-art techniques show that our proposed method improves over the benchmarks on popular non-intrusive load monitoring datasets.

Additional Key Words and Phrases: representation learning, multi-label classification, non-intrusive load monitoring.

1 INTRODUCTION

In a recent study we introduced multi-label consistent convolutional transform learning [18]. This shallow technique showed promising results in non-intrusive load monitoring (NILM). This letter proposes two major improvements over the former [18]. First, we introduce a deeper version of the representation learning model, called multi-label deep convolutional transform learning. Second, our approach can operate in both classification and regression scenarios. In classification, it detects the state of appliances (ON / OFF) while, in regression, it estimates the power consumption of individual appliances. In contrast, the former shallow formulation [18] was limited to classification, i.e. it could detect the states of appliances but could not estimate their energy consumptions. Lately, similar convolutional transform learning based frameworks have been used for unsupervised [6, 7] and supervised feature extraction [8]. In future, we would like to extend the proposed work for clustering as done in a previous transform learning based paper [17].

NILM techniques can be broadly segregated into two categories: (i) techniques that estimate the power consumption of appliances, and (ii) techniques that detect the states of appliances. Traditionally, NILM was addressed by the first class of techniques, using methods based on factorial

hidden Markov model (FHMM) [14], sparse coding (SC) [13] and combinatorial optimization [2], to estimate the power consumptions of individual appliances. The second category tackles NILM as a multi-label classification problem. A review of traditional methods can be found in [20]. More recently, off-the-shelf deep learning methods like autoencoders and long short-term memory (LSTM) networks [11], deep learning architectures based on dictionary and transform learning [19], random forests [21] or convolutional neural network (CNN) architecture [22] were also proposed for detecting the states of appliances. Up to our knowledge, most of the studies deploy some off-the-shelf algorithm with slight modifications for the multi-label setting to address NILM. In this paper, our contribution lies in proposing a method able to tackle both regression and classification problems arising in NILM, by relying on a novel representation learning paradigm. In a prior study, we introduced the framework of unsupervised convolutional transform learning [16]. In [18], we show how to integrate a multi-label cost in it so as to make the method supervised. This present work lies on a deeper version of the former model. However, unlike in [18], which could only classify, and hence detect the states of the appliances, we the proposed solution here can estimate the energy consumptions as well.

2 BACKGROUND

In this section, we introduce the basis for convolutional transform learning [18] and multi-label convolutional transform learning. These will be required for understanding our proposed multi-label deep convolutional transform learning.

2.1 Convolutional Transform Learning

Convolutional transform learning analyses some data $s^{(k)}$ ($k = 1, \ldots, K$) using a set of learnt filters $t^{(m)}$ ($m = 1, \ldots, M$) to extract a set of features $x_m^{(k)}$. The representation model reads:

$$t_m * s^{(k)} = x_m^{(k)}, \text{ for all } m \in \{1, \ldots, M\} \text{ and } k \in \{1, \ldots, K\}, \tag{1}$$

with $*$ a given discrete convolution operation with suitable dimension (typically, 1D or 2D) and padding (typically, zero). In the training stage, the convolutional filters and the associated coefficients are learnt from the data by solving the optimization problem,

$$\operatorname*{minimize}_{(t_m)_m, (x_m^{(k)})_{m,k}} \frac{1}{2} \sum_{k=1}^{K} \sum_{m=1}^{M} \left(\left\| t_m * s^{(k)} - x_m^{(k)} \right\|_2^2 + \psi\left(x_m^{(k)}\right) \right) + \mu \|T\|_F^2 - \lambda \log \det(T). \tag{2}$$

The term ψ is a suitable penalization on the features, whereas the hybrid term $\mu \|T\|_F^2 - \lambda \log \det(T)$ enforces the uniqueness of the learnt filters for some positive weights (μ, λ). This is one major difference with CNN, where such constraint would be difficult to impose during the training phase. In matrix notation, (2) can be expressed as

$$\operatorname*{minimize}_{T, X} \ F(T, X) = \frac{1}{2} \|T \star S - X\|_F^2 + \Psi(X) + \mu \|T\|_F^2 - \lambda \log \det(T), \tag{3}$$

where $T = \begin{bmatrix} t_1 & \ldots & t_M \end{bmatrix}$, $S = \begin{bmatrix} s^{(1)} & \ldots & s^{(K)} \end{bmatrix}^\top$, $X = \begin{bmatrix} x_1^{(k)} & \ldots & x_M^{(k)} \end{bmatrix}_{1 \le k \le K}$,

$$T \star S = \begin{bmatrix} t_1 * s^{(1)} & \ldots & t_M * s^{(1)} \\ \vdots & \ddots & \vdots \\ t_1 * s^{(K)} & \ldots & t_M * s^{(K)} \end{bmatrix}, \tag{4}$$

and Ψ amounts to applying the penalty term ψ column-wise on X and summing. The problem (3) can be solved using the alternating proximal algorithm [3].
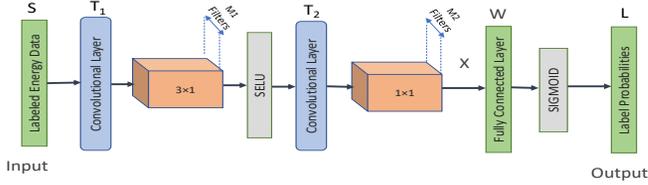
Fig. 1. Multi-label Deep Convolutional Transform Learning Architecture.

2.2 Multi-label Convolutional Transform Learning

The idea of label consistency was initially introduced for dictionary learning [10]. Later it was incorporated within the transform learning framework [15]. The label consistency cost was incorporated into the convolutional transform learning framework in [18]. Following the recent success of the label consistency term, we propose to employ it to devise a supervised version of our convolutional transform learning method. This amounts to adding the extra label consistency term in (3), leading to the resolution of:

$$\underset{T,X,M}{\text{minimize}} \quad \frac{1}{2}\|T \star S - X\|_F^2 + \Psi(X) + \mu\|T\|_F^2 - \lambda \log \det(T) + \eta\|Q - MX\|_F^2. \tag{5}$$

Here the term $\|Q - MX\|_F^2$, associated to the positive weight η, corresponds to label consistency. As in [18], matrix Q gathers the one hot encoded class labels, and M is a mapping from the learnt representation to the labels. To solve (5), one could adopt the alternating proximal algorithm [3].

3 PROPOSED MULTI-LABEL DEEP CONVOLUTIONAL TRANSFORM LEARNING

We propose a deeper extension of multi-label convolutional transform learning with a changed cost for the multi-label consistency term. Our justification for a deep architecture relies on the key property that the solution $\widehat{X}$ to (3), assuming fixed filters T, can be reformulated as the simple application of an element-wise activation function. That is:

$$\text{argmin}_X \ F(T,X) = \Phi(T \star S), \tag{6}$$

with Φ being the proximity operator of Ψ [4]. It is interesting to remark that, if Ψ is the indicator function of the positive orthant, then Φ identifies with the famous rectified linear unit (ReLU) activation function. Many other examples of mapping between proximity operators and activation functions are provided in [4]. Consequently, we propose to compute deep features by stacking several such layers, leading to $X_\ell = \Phi_\ell(T_\ell \star X_{\ell-1})$ with $\ell = 1, \ldots, L-1$ and $X_0 = S$.

For both classification and regression tasks, the input remains the same, namely the power consumption over a period of time. For classification task, the labels associated to the data S are gathered into a matrix L, where each column is a binary vector with the n-th element being 0 if the n-th appliance is off and 1 if the n-th appliance in on. Owing to such binary nature, it is more

appropriate to use a binary cross entropy loss for label consistency, leading to:

$$\underset{T,X,W}{\text{minimize}} \quad \frac{1}{2} \|T_2 \star \Phi(T_1 \star S) - X\|_F^2 + \Psi(X) + \sum_{\ell=1}^{2} \left(\mu \|T_\ell\|_F^2 - \lambda \log \det (T_\ell) \right) + \eta J_{BCE}(\sigma(WX), L), \quad (7)$$

where σ is the sigmoid function, and J_{BCE} is the binary cross-entropy loss. Instead of predicting the state of the appliance, if we want to predict the power consumption, the labels in the matrix L will consist of appliance-wise power consumption. Since the labels will be real values, we use the Euclidean cost, which yields:

$$\underset{T,X,W}{\text{minimize}} \quad \frac{1}{2} \|T_2 \star \Phi(T_1 \star S) - X\|_F^2 + \Psi(X) + \sum_{\ell=1}^{2} \left(\mu \|T_\ell\|_F^2 - \lambda \log \det (T_\ell) \right) + \eta \|WX - L\|_F^2. \quad (8)$$

Hereabove, we show the formulations for two layers of convolutional transforms (T_1 and T_2), but it can be extended to more in a straightforward way. We finally propose to solve both problems (7) and (8) using backpropagation with accelerated gradient descent [12]. Fig. 1 shows the schematic diagram of our proposed method. While it appears to be similar to that of a convolutional neural network (CNN), the key difference of our proposed approach lies in the way the convolutional filters are learnt. Here, we guarantee uniqueness of the learnt filters, while CNN does not. The later start with random initialization of each filter and 'hopes' that the filters will be unique.

4 EXPERIMENTAL EVALUATION

We have experimented on two popular NILM datasets – REDD and Pecan Street. Both are available in NILM Tool Kit. Owing to limitations in space, we cannot describe the datasets in detail. For both datasets, high frequency data is available. To emulate real-world conditions, we have aggregated and sub-sampled the data to one sample per minute. For both datasets, 60 % of the houses have been used for training and the rest for of testing.

4.1 Classification

In the multi-label classification scenario, we have compared with two state-of-the-art techniques, namely deep learning based NILM with pinball loss (PB-NILM) [5] and multi-label deep transform learning (MLDTL) [19]. For our proposed technique the parameters were determined using k-fold cross validation on the training data. The retained parameters were $\beta = 1$, $\mu = 3$, $\lambda = 1$ and $\eta = 1$. The $F1_{macro}$, the $F1_{micro}$ and the average energy error (AEE), that are standard measures for multi-label classification based NILM [20], are presented in Tab. I and Tab. II.

For both the datasets, we see that our proposed algorithm with two layers performs the best. Adding further layers on these small datasets results in overfitting. We find that PB-NILM works better for Pecan Street (larger dataset) compared to REDD which may owe to the fact that the technique is over-fitting for the smaller data.

4.2 Regression

In this scenario, our objective is to predict the energy consumed by different appliances. We have compared against deep latent generative model (DLGM) [1] and semi-binary non-negative matrix factorization (SMNNMF) [9]. The parameters for the existing benchmarks have been obtained from the papers. The parametric values for our model remain the same as before. For comparing the accuracies, we compute the normalized energy errors of common appliances for the two different datasets. The results are shown in Tables III and IV. Here we are showing the best results from our two layer architecture.

Table 1. **Classification Results on REDD**

Method	Macro F1 Measure	Micro F1 Measure	Average Energy Error
PB-NILM	0.5515	0.5576	0.3903
ML-DTL	0.5693	0.5642	0.3537
Proposed 1 layer	0.5687	0.5682	0.2926
Proposed 2 layer	**0.6018**	**0.6026**	**0.2558**
Proposed 3 layer	0.5425	0.5419	0.3282

Table 2. **Classification Results on Pecan Street**

Method	Macro F1 Measure	Micro F1 Measure	Average Energy Error
PB-NILM	0.6231	0.6207	0.2582
ML-DTL	0.5552	0.5552	0.4048
Proposed 1 layer	0.6121	0.6119	0.2723
Proposed 2 layer	**0.6381**	**0.6378**	**0.2316**
Proposed 3 layer	0.5983	0.5963	0.2902

5 CONCLUSION

This work proposes a new supervised deep learning framework. It is based on the concept of convolutional transform learning. We specified this framework to the task of non-intrusive load monitoring. It can tackle both to classification problem for identifying the states of the appliances and to the regression problem for estimating their energy consumptions. Comparisons with state-of-the-art techniques show that our proposed method improves over the rest. We expect that the results can be further improved by adopting post-processing approaches such as [9].

REFERENCES

[1] Gissella Bejarano, David DeFazio, and Arti Ramesh. 2019. Deep Latent Generative Models for Energy Disaggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 850–857.

[2] Md Zulfiquar Ali Bhotto, Stephen Makonin, and Ivan V Bajić. 2016. Load disaggregation based on aided linear integer programming. *IEEE Transactions on Circuits and Systems II: Express Briefs* 64, 7 (2016), 792–796.

[3] Emilie Chouzenoux, Jean-Christophe Pesquet, and Audrey Repetti. 2016. A block coordinate variable metric forward–backward algorithm. *Journal of Global Optimization* 66, 3 (2016), 457–485.

[4] P.L. Combettes and J.-C. Pesquet. 2018. Deep Neural Network Structures Solving Variational Inequalities. *Set-Valued and Variational Analysis (2018)* (2018). https://arxiv.org/abs/1808.07526

[5] Eduardo Gomes and Lucas Pereira. 2020. PB-NILM: Pinball guided deep non-intrusive load monitoring. *IEEE Access* 8 (2020), 48386–48398.

[6] Pooja Gupta, Jyoti Maggu, Angshul Majumdar, Emilie Chouzenoux, and Giovanni Chierchia. 2020. ConFuse: Convolutional Transform Learning Fusion Framework For Multi-Channel Data Analysis. *arXiv preprint arXiv:2011.04317* (2020).

[7] Pooja Gupta, Jyoti Maggu, Angshul Majumdar, Emilie Chouzenoux, and Giovanni Chierchia. 2020. DeConFuse: a deep convolutional transform-based unsupervised fusion framework. *EURASIP Journal on Advances in Signal Processing* 2020, 1 (2020), 1–32.

[8] Pooja Gupta, Angshul Majumdar, Emilie Chouzenoux, and Giovanni Chierchia. 2020. SuperDeConFuse: A supervised deep convolutional transform based fusion framework for financial trading systems. *Expert Systems with Applications* (2020), 114206.

[9] Kanghang He, Dusan Jakovetic, Bochao Zhao, Vladimir Stankovic, Lina Stankovic, and Samuel Cheng. 2019. A generic optimisation-based approach for improving non-intrusive load monitoring. *IEEE Transactions on Smart Grid* 10, 6

(2019), 6472–6480.

[10] Zhuolin Jiang, Zhe Lin, and Larry S Davis. 2013. Label consistent K-SVD: Learning a discriminative dictionary for recognition. *IEEE transactions on pattern analysis and machine intelligence* 35, 11 (2013), 2651–2664.

[11] 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.

[12] D. P. Kingma and J. Ba. 2015. Adam: A method for stochastic optimization. *Proc. of ICLR (2015)* (2015). https://arxiv.org/abs/1412.6980

[13] J Zico Kolter, Siddharth Batra, and Andrew Ng. 2010. Energy disaggregation via discriminative sparse coding. *Advances in neural information processing systems* 23 (2010), 1153–1161.

[14] J Zico Kolter and Tommi Jaakkola. 2012. Approximate inference in additive factorial hmms with application to energy disaggregation. In *Artificial intelligence and statistics*. PMLR, 1472–1482.

[15] Jyoti Maggu, Hemant K Aggarwal, and Angshul Majumdar. 2019. Label-consistent transform learning for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters* 16, 9 (2019), 1502–1506.

[16] Jyoti Maggu, Emilie Chouzenoux, Giovanni Chierchia, and Angshul Majumdar. 2018. Convolutional transform learning. In *International Conference on Neural Information Processing*. Springer, 162–174.

[17] Jyoti Maggu, Angshul Majumdar, Emilie Chouzenoux, and Giovanni Chierchia. 2020. Deeply Transformed Subspace Clustering. *Signal Processing* (2020), 107628.

[18] Shikha Singh, Jyoti Maggu, Angshul Majumdar, Emilie Chouzenoux, and Giovanni Chierchia. 2020. Multi-Label Consistent Convolutional Transform Learning: Application to Non-Intrusive Load Monitoring. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4287–4291.

[19] Vanika Singhal, Jyoti Maggu, and Angshul Majumdar. 2018. Simultaneous detection of multiple appliances from smart-meter measurements via multi-label consistent deep dictionary learning and deep transform learning. *IEEE Transactions on Smart Grid* 10, 3 (2018), 2969–2978.

[20] Seyed Mostafa Tabatabaei, Scott Dick, and Wilsun Xu. 2016. Toward non-intrusive load monitoring via multi-label classification. *IEEE Transactions on Smart Grid* 8, 1 (2016), 26–40.

[21] Xin Wu, Yuchen Gao, and Dian Jiao. 2019. Multi-label classification based on random forest algorithm for non-intrusive load monitoring system. *Processes* 7, 6 (2019), 337.

[22] Yandong Yang, Jing Zhong, Wei Li, T Aaron Gulliver, and Shufang Li. 2019. Semi-Supervised Multi-Label Deep Learning based Non-intrusive Load Monitoring in Smart Grids. *IEEE Transactions on Industrial Informatics* (2019).