# An Empirical Study of Ladder Network and Multitask Learning on Energy Disaggregation in Taiwan

1st Fang-Yi Chang
*Digital Transformation Institute*
*Institute For Information Industry*
Taipei, Taiwan
fangyichang@iii.org.tw

1nd Chun Chen
*Department of Computer Science and Information Engineering*
*National Taiwan University*
Taipei, Taiwan
b03902033@ntu.edu.tw

2rd Shou-De Lin
*Department of Computer Science and Information Engineering*
*National Taiwan University*
Taipei, Taiwan
sdlin@csie.ntu.edu.tw

***Abstract*—Energy disaggregation is a technique of estimation electricity consumption of individual appliance from an aggregated meter. In this paper, we study ladder network [6] and multitask learning on energy disaggregation using auto-encoder architecture. This auto-encoder architecture was proposed from Kelly and Knottenbelt in their recent research work [1]. We used this auto-encoder architecture to the high-ownership appliances, air conditioner, bottle warmer, fridge, television and washing machine, in Taiwan and evaluated the effectiveness of the ladder network and multitask learning via these five appliances. The experimental data set has collected by Institute For Information Industry. We expect that this project can promote the industrial development of big data-driven smart energy management in Taiwan.**

***Index Terms*—Energy Disaggregation; Deep Learning; Multitask Learning; Semi-supervised; Smart Meter; Value-Add Electricity Services; Energy Saving**

## I. Introduction

Energy disaggregation is a technique to infer an appliance electricity consumption individually from a single meter that measures the electricity consumption of multiple appliances. One of the applications in the field is under the framework of the ongoing smart grid system with the goal to help end users to reduce energy consumption. Research on energy disaggregation started with the works of Hart in the mid-1980s [3][4], and these works were to extract transitions between steady-states. Recently, Kelly and Knottenbelt [1] applied deep learning to energy disaggregation and have shown that auto-encoder architecture outperformed than either combinatorial optimization or factorial hidden Markov model. Deep learning now is a dominant approach for its significant achievement in many fields, including computer vision, speech recognition and natural language process. Based on the achievement, we investigated whether the other deep learning techniques, semi-supervised learning and multitask learning, in machine learning can be applied to energy disaggregation using auto-encoder architecture.

We thank the Bureau of Energy, Ministry of Economic Affairs of Taiwan, ROC for the financial support under Contract No.107-E0215.

In energy disaggregation, since we have much access to much more unlabelled data than labelled data [1], it is interesting to both labelled and unlabelled data via semi-supervised learning. In deep learning, semi-supervised usually refers to learning a representation so that the observations from the similar condition have similar representations, and there are many semi-supervised algorithms to be developed. In this paper, since we use the auto-encoder architecture to perform energy disaggregation, we apply the ladder networks [6], which structure is an auto-encoder with skip connections from the encoder to decoder, and the learning task is similar to that in denoising autoencoders but applied to every layer, nut only the inputs. On the other hand, multitask learning can improve generalization because of the shared parameters via leveraging the relatedness information contained in the training signals of related tasks [2]. When there are relatedness between individual tasks, it can be advantageous to learn all task at the same time than learning each individual task. We except that these two techniques can increase the generalization in the field of energy disaggregation.

Our main contributions are the following:

- Investigate whether the auto-encoder architecture proposed by Kelly and Knottenbelt [1] is effective on the common five appliances, air conditioner, bottle warmer, fridge, television and washing machine, in Taiwan. These five appliances have significant meaning in Taiwan for their high number of ownership and their long usage time

2376-6824/18/$31.00 ©2018 IEEE
DOI 10.1109/TAAI.2018.00028

[11]. The latest statistics figure is given on Table II.

- Investigate whether the two technique, the ladder network and multitask learning, are effective on energy disaggregation. We evaluate how well these two techniques improve the generalization in the unseen houses via the common metrics in time-series data and energy disaggreagtion.

This paper is structured as follows: In Section II we briefly introduce to the ladder networks and multitask learning. In section III we describe the auto-encoder architecture from Kelly and Knottenbelt [1]. In section IV we present the empirical results for the single-task learning, the ladder networks and multitask learning. Finally, in section V we discuss our results and offer our suggestions for future research.

Table I
OVERVIEW OF THE AVERAGE STATISTICS FIGURES FOR THE FIVE APPLIANCES IN TAIWAN [11]

| Appliance | Ownership Number | Usage Time (hour-minute) |
|---|---|---|
| air conditioner | 2.7 | 1-06 |
| bottle warmer | 0.5 | 10-48 |
| fridge | 1.1 | 23-54 |
| television | 1.3 | 4-54 |
| washing machine | 1.0 | 0-48 |

## II. INTRODUCTION TO THE LADDER NETWORKS AND MULTITASK LEARNING

### A. the ladder networks

The objective of semi-supervised learning is to learn a mapping that models the underlying distribution via both labeled and unlabeled data. The mapping in the ladder network is a deep denoising auto-encoder in which noise distributed as Gaussian distribution is injected into all hidden layers, and the cost function is a weighted sum of the supervised Cross Entropy on the top of the encoder and the unsupervised denoising square error costs at each layer of the decoder. Since all layers are corrupted by noise, another encoder with shared parameters is to provide the noiseless reconstruction targets.

The design of lateral skip connection in the ladder network allows each layer of the noisy encoder to connect to its corresponding layer in the decoder. Each layer of the decoder has two sources of signal, one from the above layer and other from the corresponding layer in the encoder. This design enables the higher layers to remain more abstract. In the formal language, the ladder network and its cost function are defined as follows:

$$\tilde{x}, \tilde{z}^{(1)}, ..., \tilde{z}^{(L)}, \tilde{y} = Encoder_{noisy}(x) \tag{1}$$

$$x, z^{(1)}, ..., z^{(L)}, y = Encoder_{clean}(x) \tag{2}$$

$$\hat{x}, \hat{z}^{(1)}, ..., \hat{z}^{(L)} = Decoder(\tilde{z}^{(1)}, ..., \tilde{z}^{(L)}) \tag{3}$$

$$Cost = -\sum_{i=1}^{N} logP(y^*(i) = \tilde{y}(i)|x(i)) + \sum_{j=N+1}^{M}\sum_{l=1}^{L}\lambda_l||z^{(l)} - \hat{z}^{(l)}||_2^2 \tag{4}$$

where the variables x, y, $\tilde{y}$ and $y^*$ are the input, the noiseless output, the noisy output the true labeled value. The variables $z^{(l)}$, $\tilde{z}^{(l)}$ and $\hat{z}^{(l)}$ are hidden representation, its noisy version and its reconstructed version at layer l. L is the number of layers, and N, M are the number of labeled data and unlabeled data, respectively. In this paper, since our experimental data set is time-series, the cost function is need to adjust to (5).

$$Cost = -\sum_{i=1}^{N} ||y^*(i) - \tilde{y}(i)||_2^2 + \sum_{j=N+1}^{M}\sum_{l=1}^{L}\lambda_l||z^{(l)} - \hat{z}^{(l)}||_2^2 \tag{5}$$

### B. multitask learning

The motivation for multitask learning is that human can apply the ‘similarity’ that acquired from the related tasks in the past to a new task. The similarities between many tasks human learn are what enable human to learn new task [2]. For example, a baby can recognize human faces and then can recognize other objects. These two task are not the same but related. In machine learning, multitask learning is effective because of its statistical data augmentation, eavesdropping and representation bias.

- Statistical data augmentation: related tasks may have independent noise source but have a common hidden layer feature. Learning the related tasks simultaneously can leveraging the different source of noise to learn a hidden layer feature more generally than learning individual tasks.
- Eavesdropping: consider that a hidden features are easy to learn for some tasks but difficult for another tasks. This may be either because that the difficult tasks interact with the hidden features in a more complex way during the training process, or because that other features within the tasks hinder the ability of net to learn the hidden feature. Multitask learning can help the difficult ones to eavesdrop the hidden feature for the easy ones.
- Representation bias: each net individually learning the tasks that are related may have multiple local minima, and if a net simultaneously learning these related tasks, the net tend to search the common local minima in the representative space. This leads the net to use the hidden layer representation that can be used by more tasks.

In energy disaggregation, the tasks of disaggreation for each appliance are related. For example, all target appliances within the same time interval have the same baseline that is sum from other appliances except for the target ones. However, the nets individually learning target appliances may find the different baselines. In our experiment, we apply multitask learning to disaggregation and expect that it can enhance the accuracy via leveraging the relatedness.

## III. Architecture

### A. single Task

We perform these following architectures:

- Input with sequence length 60
- 1D conv with filter size=4, stride=1, number of filters=8
- Fully connected with 456 units
- Fully connected with 128 units
- Fully connected with 456 units
- 1D conv with filter size=4, stride=1, number of filters=1
- Output with sequence length 60

The activation functions for all 1D convolutions and all fully-connected layers are ReLU and linear, respectively. We investigate the effectiveness of convolution operations in supervised, ladder network and multitask in our five target appliances. In section IV, we use CNN_DNN to refer the whole above architecture and DNN to refer the fully-connected parts of the whole architecture.

### B. the ladder network task and multitask learning

For the ladder netowrk, We use the same architectures as single task but the activation function for the output layer is softmax. For the CNN_DNN, we modified our models following the paper [12]. Noisy is injected into the convolutional layers in encoder, and the corresponding decoder are deconvolutional layers. For the multitask learning, we train the five tasks at one time, and try the possible composition of shared layers and task-specifics layers based on the same architectures as single task. Since the space is limited, we only present the best result in the following section.

## IV. Experiment

### A. III dataset

The raw data is collected by Institute For Information Industry which consists of near 30 different households from Northern Taiwan about six months. For each household, we recorded (1) the aggregated electricity consumption (Watt) with sample rate 1 minute; (2) up to 1 the appliance's electricity consumption for each category (the five appliances) with sample rate 1 minute.

### B. evaluation metrics

We use two metrics to evaluate our models: mean absolute error and relative error in total energy.

$$\text{mean absolute error} = \frac{1}{T}\sum_{t=1}^{Y} |\hat{y}_t - y_t| \tag{6}$$

$$\text{relative error in total energy} = \frac{\hat{E} - E}{max(\hat{E}, E)} \tag{7}$$

where y and $\hat{y}$ are the actual value and the estimated value, respectively, at time point t. E and $\hat{E}$ are the total actual value and the total estimated value, respectively. The units of y, $\hat{y}$, E and $\hat{E}$ are Watt.

Mean absolute error are the common metrics in time-series data. When energy disaggregation is used in the field, the accuracy of estimation of the total consumption within intervals is quite important. Thus, we include relative error in total energy into our evaluation.

### C. empirical result

We use the same training data set and unseen data set for all experiments in single task, the ladder network and multitask. The test result of unseen data are in Table III-VIII. Please note that when the REITE (relative error in the total energy) is 1 or near 1, it means that the corresponding models failed to learn the underlying distribution and their all neural weights were near zero, though other statistics numbers, mean absolute error are better. This is because that the usage times of the three appliances, air conditioner, television and washing machine, are averagely smaller than 5 hours according to Table II. This means that the electricity consumption of these three appliances remain near zero value most time, thus the failed models have lower mean squared error and mean absolute error. However, the failed models cannot detect or infer the electricity consumption when these appliances are on.

In single task, the auto-encoder architecture is effective on the five common appliances in Taiwan. We can see the effectiveness of the operation of convolution. This operation indeed increases the accuracy on our data set, since it can dramatically enhance the efficiency of edge detection [8] for its sparse connectivity and parameter sharing.

Table II
Single Task : Mean absolute error

| | AC | B | F | T | W |
|---|---|---|---|---|---|
| DNN | 47.41 | 40.24 | 39.99 | 25.40 | 1.15 |
| CNN_DNN | 25.78 | 98.99 | 37.68 | 34.36 | 2.41 |

Table III
Single Task : Relative error in total energy

| | AC | B | F | T | W |
|---|---|---|---|---|---|
| DNN | -0.80 | 1 | -0.35 | 1 | 1 |
| CNN_DNN | -0.50 | -0.40 | -0.27 | 0.54 | -0.08 |

In the ladder network, the convolution operation is not effective as it did in the single task, though it was successful in the field of Human Activity Recognition [12]. The ladder network tends to regress smooth lines at the value of average for the corresponding appliances within intervals. Beside, the ladder network is ineffective on the 'small' appliance, television, for its small consumption. However, in terms of REITE, the ladder network has performed better for bottle warmer than the single learning, and the result of REITE for fridge is accepted. It is suitable for the ladder network to be a regressor for average values in bottle warmer and fridge. The other three appliances, air conditioner, television and washing machine, are failure for their few 'on' data.

Table IV
THE LADDER NETWORK: MEAN ABSOLUTE ERROR

| | AC | B | F | T | W |
|---|---|---|---|---|---|
| DNN 3:1 | 85.32 | 73.69 | 44.65 | 97.42 | 22.85 |
| DNN1:1 | 101.11 | 80.41 | 46.58 | 100.86 | 39.89 |
| DNN 1:3 | 146.90 | 77.53 | 45.31 | 124.38 | 32.22 |
| CNN_DNN 3:1 | 14.37 | 41.94 | 38.88 | 25.94 | 1.37 |
| CNN_DNN 1:1 | 16.98 | 40.73 | 35.29 | 17.06 | 33.41 |
| CNN_DNN 1:3 | 37.18 | 51.63 | 34.69 | 27.41 | 16.87 |

topology_name_ #1:#2 means that the ratio of label and unlabeled is #1 and #2 .

Table V
THE LADDER NETWORK : RELATIVE ERROR IN TOTAL ENERGY

| | AC | B | F | T | W |
|---|---|---|---|---|---|
| DNN3:1 | -0.86 | 0.22 | -0.38 | 1 | -0.85 |
| DNN1:1 | -0.90 | 0.24 | -0.46 | 1 | -0.96 |
| DNN 1:3 | -0.93 | 0.21 | -0.45 | 1 | -0.94 |
| CNN_DNN 3:1 | 0.79 | 0.99 | 1 | 1 | 0.95 |
| CNN_DNN 1:1 | 1 | 1 | 1 | 1 | 1 |
| CNN_DNN 1:3 | 1 | 0.99 | 0.99 | 1 | -0.90 |

topology_name_ #1:#2 means that the ratio of label and unlabeled is #1 and #2 .

The multitask improves the 'small' and 'medium' appliance but the 'larger' one. The best improvement is bottle warmer and the worse is air conditioner. Overall, the mean square error and mean absolute error are improved. Compared to all the experiments in single task, most of them are improved. The combination we used in the multitask experiments is to train the five appliance at the same time. The shared layers are the layers before the fully-connected layer with 128 units(including this layer), and the specific layers is the remaining layers.

## V. FUTURE

We present the empirical results of the single task, ladder network and multitask learning for the five common appliances in Taiwan. For the single task, the auto-encoder architecture is effective; For the ladder network, it is suitable to be a regressor for average values for the 'longer' usage-time appliances; For multitask learning, the overall performance is improved. However, there is still plenty of work to do. For example, we can further investigate the effectiveness of other semi-supervised techniques, such as pre-train. We also can further try many more complicate multitask structure in the future.

Table VI
MULTITASK : MEAN ABSOLUTE ERROR

| | AC | B | F | T | W |
|---|---|---|---|---|---|
| DNN | 35.85 | 44.54 | 40.30 | 35.01 | 1.15 |
| CNN_DNN | 42.58 | 45.43 | 37.15 | 33.85 | 5.10 |

Table VII
MULTITASK : RELATIVE ERROR IN TOTAL ENERGY

| | AC | B | F | T | W |
|---|---|---|---|---|---|
| DNN | -0.72 | 0.76 | -0.26 | 0.38 | 1.00 |
| CNN_DNN | -0.77 | 0.69 | -0.12 | 0.46 | -0.72 |

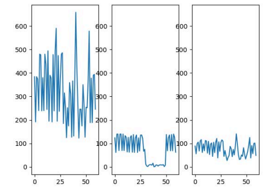

Figure 1. Example energy disaggregation from the aggregated meter in an unseen house. The left is the aggregated meter, the middle is the true fridge's value, and the right is the estimated fridge's value.


## ACKNOWLEDGMENT

We thank the Bureau of Energy, Ministry of Economic Affairs of Taiwan, ROC for the financial support under Contract No.107-E0215, Yung-Chieh Hung and Kuei-Chun Chiang for their assistance in the project, and Hao-Ying Liang for their assistance in Keras and Tensorflow.


## REFERENCES


[1] Jack Kelly and William Knottenbelt. *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 (BuildSys 15). ACM, New York, NY, USA, 5564, 2015.

[2] Rich Caruana. *Multitask Learning*. Thesis, MIT Energy Laboratory and Electric Power Research Institute, 1997.

[3] G. W. Hart. *Prototype nonintrusive appliance load monitor*. Technical report, School of Computer Science Carnegie Mellon University, 1985.

[4] G. W. Hart. *Nonintrusive appliance load monitoring*. Proceedings of the IEEE, 80(12):18701891, 1992.

[5] Nikolaos Chrysogelos. *Energy Disaggregation via Current Smart Metering Infrastructure - Study and Implementation Based on Present Constraints*. aSemester Thesis, Power Systems Laboratory, ETH Zurich, 2016.

[6] Antti Rasmus and Harri Valpola and Mikko Honkala and Mathias Berglund and Tapani Raiko. *Semi-Supervised Learning with Ladder Networks*. arXiv preprint arXiv:1507.02672, 2015.

[7] Mohammad Pezeshki and Linxi Fan and Philemon Brakel and Aaron Courville and Yoshua Bengio. *Deconstructing the Ladder Network Architecture*. arXiv preprint arXiv:1511.06430, 2015.

[8] Lan Goodfellowm Yoshua Bengio and Aaron Courvilleo. *Deep Learning*. Massachusetts Institute of Technology, 2016.

[9] Kolter, J.Z. and Johnson, M.J. *REDD : A Public Data Set for Energy Disaggregation Research*. MProceedings of the SustKDD Workshop on Data Mining Applications in Sustainability, San Diego, CA, USA, 2011.

[10] Jack Kelly and William Knottenbelt. *The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes*. 2016.

[11] Su-Chin Lin and Allen Lin. *The Residential Sector Electricity Use Research in Taiwan*.(Chinese) Journal of Taiwan Energy Volume 4, No. 3, 2017.

[12] Ming Zeng and Tong Yu and Xiao Wang and Le T. Nguyen and Ole J. Mengshoel and Ian Lane. *Semi-Supervised Convolutional Neural Networks for Human Activity Recognition*. arXiv preprint arXiv:1801.07827, 2018.