Non-intrusive Load Monitoring Using Imaging Time Series and Convolutional Neural Networks

Mohammad Mottahedi[1], Somayeh Asadi[2]

1) Ph.D. Student, Department of Architectural Engineering, The Pennsylvania State University, University Park, Pennsylvania, USA. Email: m.mottahedi@psu.edu
2) Assistant Professor, Department of Architectural Engineering, The Pennsylvania State University, University Park, Pennsylvania, USA. Email: asadi@engr.psu.edu

Abstract:

In recent years, more than 50 million advanced (smart) metering infrastructure units have been installed by the U.S electric utilities. Although, smart metering can provide hourly or sub-hourly customer load, it has failed to directly benefit and provide actionable information to consumers and engage them in energy savings. Using non-intrusive load monitoring techniques, the smart metering data can be disaggregated to individual components for each appliance which consequently can be used to monitor the performance of the appliances, inform consumer about future failures and help them to reduce energy consumption. In this study, the time series data was encoded to images using Gramian Angular Summation Fields (GASF). This enabled us to train Convolutional Neural Networks (CNNs) on the encoded images. The developed model was used for energy disaggregation. To determine the performance of the developed model, mean absolute error and relative error in total energy was calculated for the obtained results from the deep neural net architecture. The convolutional neural network generalized well on unused test data without significant effort in data cleaning and feature engineering.

Keywords: Non-intrusive load monitoring, Energy disaggregation, Deep neural networks, Geramian Angular Summation, Computer Vision, Convolutional Neural Networks.

1. INTRODUCTION

In the United States, the residential building account for as much as 38% of electricity consumption (Franco, 2008). It is estimated that close to 8% of the electricity demand by residential building could be avoided with little or no cost (Carrie Armel; Gupta et al., 2013). In recent years, more than 50 million smart meters were deployed across the United States and the billions of dollars spent to install them in order to benefit the consumers directly with actionable information about the devices and the systems in an average household. Numerous studies demonstrated that the feedback provided by energy monitors will enable the consumers to automatically reduce their consumption by 5-15% by triggering the energy conserving behavior (Darby, 2008). One of the techniques that can provide information to electricity consumers from smart meter data is Non-intrusive load monitoring (NILM). Nonintrusive load monitoring is a technique for dissecting the aggregate electricity signal from the smart meters into individual appliance and systems that use electricity.

In order to develop feasible solution, the energy disaggregation techniques should be economically viable, without the need for additional cost for hardware and installation. Some of the challenges in developing such a viable solution to load disaggregation include using low resolution smart meter signal (15 minutes or hourly signal) and having a large variety of household appliances. In addition, most previously proposed solutions heavily rely on hand-engineered data which is time consuming and potentially unreliable. Another limitation in previous studies on non-intrusive load monitoring is the lack of large dataset for training and testing (Berges; Goldman et al., 2010; Kelly & Knottenbelt, 2015). Many of these challenges could be addressed by developing sophisticated machine learning algorithms tailored for electricity disaggregation. In recent years, deep learning became one of the most effective approaches in image classification, speech recognition and natural language processing. LeCun have shown that using Convolutional Neural Networks (CNNs) better pattern recondition can be achieved using automatic learning compared to hand-designed heuristics (Lecun; Bottou et al., 1998). CNNs also proved to be effective in capturing the translational invariance (Sainath; Mohamed et al., 2013). The advances in these fields can be adopted to improve the nonintrusive load monitoring accuracy and effectiveness.

In this study, the time series data was encoded to images using Gramian Angular Summation Fields (GASF). This enabled us to train CNNs on the encoded images. To determine the performance of the developed model, mean absolute error and relative error in total energy was calculated for the obtained results from the deep neural net architecture. The developed model will be used to provide information the electricity consumption of the appliances which can lead to reduction in electricity consumption.

2. METHODS

2.1 Non-Intrusive Load Monitoring Dataset

In this study we used Dataport (Parson; Fisher et al., 2015) dataset created by Pecan Street Inc, and further integrated to the NILMTK (Batra; Kelly et al., 2014) Software package which was developed to support energy disaggregation research. This dataset consisted of 1 minute circuit-level and building level electricity data for 669 household for duration of one month. The 669 household data was divided into two parts, the first part was used for training the neural nets and the second part (i.e. 10% of data) was left for validation.

Four appliances were chosen for this study: kettle, microwave, air conditioner, and fridge. The appliance activation signal was extracted using the NILMTK's get.activation () function by providing some threshold value for power and duration which any activation shorter than the threshold value were ignored. Training data was randomly generated (50% probability) based on the following rules: either include activation signal in the training sample or use window of time without any activation signal (Kelly & Knottenbelt, 2015).

2.2 Imaging Time Series

To exploit the recent achievements in supervised and unsupervised learning, specially convolutional neural nets which was successful in computer vision, the electricity signal was encoded into images using approach proposed by (Wang & Oates, 2015). Encoding the time series data to images will allow the learning algorithm to learn the image structures and patterns.

2.2.1 ***Gramian Angular Summation Field***

Gramian Angular Summation Field (GASF) is a polar representation of the time series (Wang & Oates, 2015).The matrix representing an image is the cosine of the summation of the angles. The time series $X = \{x_1, x_2, x_3, \dots, x_n\}$ with n real valued observation was rescaled to $[0,1]$ using equation (1):

$$\widetilde{x}_i = \frac{x_i - min(X)}{max(X) - min(X)} \tag{1}$$

The rescaled time series $\tilde{X}$ can be represented in polar coordinates using angular cosine ϕ and time stamps t_i as the radius using equations (2 and 3):

$$\phi = arccos(\tilde{x}_i), -1 < \tilde{x} < 1, \tilde{x}_i \in \widetilde{X} \tag{2}$$

$$r = \frac{t_i}{N}, t_i \in \mathbb{N} \tag{3}$$

where N is a constant factor to regularize the span of coordinate system.

The GASF representation used in this study (Wang & Oates, 2015) has the advantage that the transformed times series is bijective (has one-to-one correspondence) since cosine is monotonic in $[0, \pi]$. In addition the GASF mapping has a unique representation in polar coordinates and preserve the absolute temporal relations (Wang & Oates, 2015).

The GASF can be calculated using equations (4 and 5):

$$GASF = \begin{bmatrix} \cos(\phi_1 + \phi_1) & \cdots & \cos(\phi_1 + \phi_n) \\ \vdots & \ddots & \vdots \\ \cos(\phi_n + \phi_1) & \cdots & \cos(\phi_n + \phi_n) \end{bmatrix} \tag{4}$$

$$GASF = \widetilde{X}' \cdot \tilde{X} - \sqrt{I - \tilde{X}^2}' \cdot \sqrt{I - \tilde{X}^2} \tag{5}$$

where I is unit row vector.

The main diagonal of the GASF contains the original value information and can be used to reconstruct the time series from high level featured learned by convolutional neural net. The representative plot illustrating the original time series from aggregate (Figure 1a), appliance activation (Figure 1b) and corresponding GASF mappings (Figures 1c and1d) can be found below.

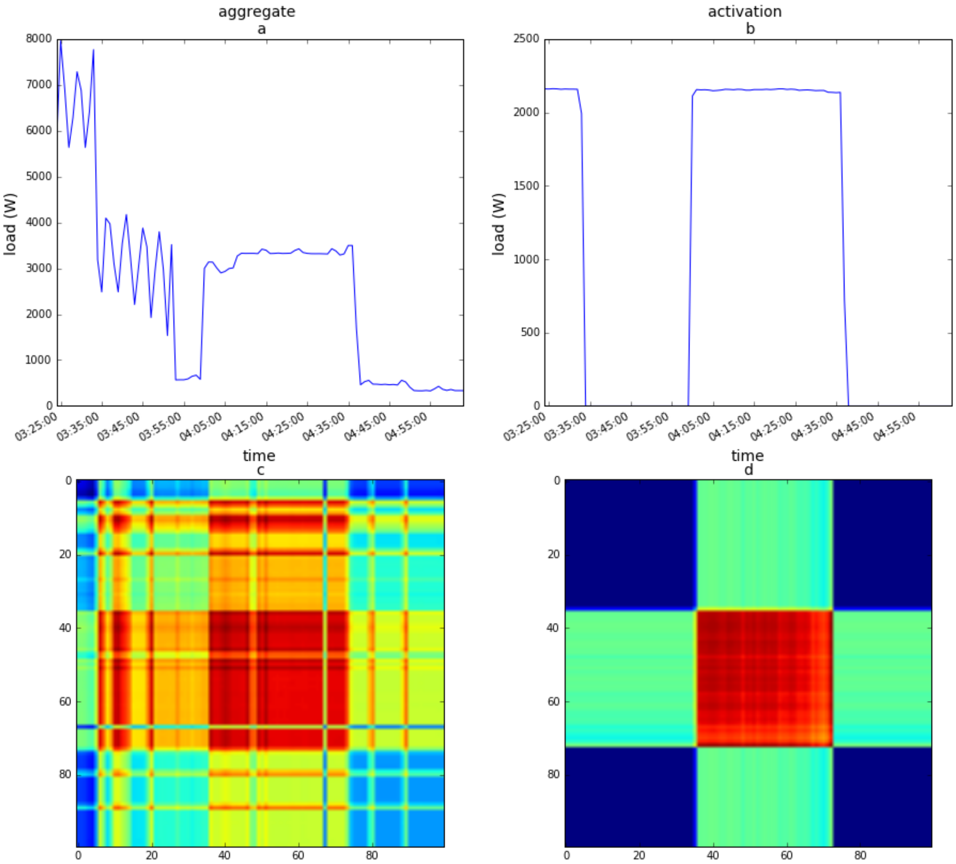

Figure 1. (a) Aggregate electricity, (b) appliance activation, (c) corresponding aggregate GASF mapping, (d) corresponding appliance activation GASF mapping.

2.3 Convolutional Neural Networks

CNNs, directly influenced from research in neuroscientific principles, are specialized type of neural networks design to process and analyze structured grid such as topology (Goodfellow; Courville et al., 2015). CNNs use at least one layer performing mathematical convolution on the input data. They have three significant improvement over traditional convolutional neural network (Goodfellow; Courville et al., 2015). First, they have sparse interaction which means less memory is needed for storing the parameters. In deep CNNs, deeper level can indirectly interact with the input which allows for representing complicated interaction by constructing interactions from simple blocks. Parameter shearing uses the same member of kernel at every position of the input which results in further reduction in storage requirement. Lastly, the equivariance to translation which means the output changes in the same way if the input changes. In addition, the convolutional neural network allows using input with variable size.In each layer of convolutional neural networks, there are two additional steps including non-linear activation function, for example a rectified linear activation, and a pooling function. The pooling function which replaces the output of the net with a summary statistic of nearby outputs results in representations that are invariant to small translations.

The convolutional neural network architecture used in this study is a variation of (Lecun; Bottou et al., 1998). The architecture includes five convolutional layer consisting convolution, a rectified linear activation and max pooling function. Each convolutional layer have 5×5 pixel kernel and the max pooling layer has 2×2 pixel size (Figure 2). In the last two layers, there are two fully connected dense layers and linear activation between them. The loss function is root mean-squared-error (RMSE) which optimized using stochastic gradient descent with learning rate and Nesterov momentum respectively set to 0.01 and 0.9.

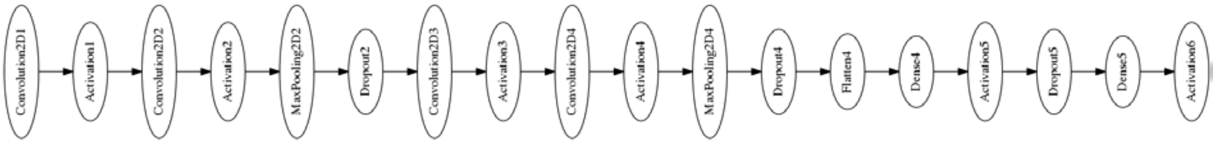

Figure 2. Deep convolutional network architecture

2.4 Disaggregation

In this study, a separate network was trained to disaggregate the electricity load for each of the three selected appliances. Depending on the nature of the appliance energy consumption pattern, a specific activation window length was selected. For example in the case of air conditioner, the activation window is 100 units equivalent to 1 hour and 40 minutes. After training the nets, the portion of the data which was set aside for testing, was shown to the net as the input. The input data was sectioned to portion with the length equal to the target appliance window. The disaggregated time series was reconstructed (Wang & Oates, 2015) from the output of the net and converted from polar representation using the main diagonal of the GASF matrix using equation (7):

$$cos(\phi) = \sqrt{\frac{cos(2\phi)+1}{2}}, \phi \in [0, \pi/2] \quad (7)$$

The resulting time series were concatenated to construct the whole disaggregate appliance load. To determine the accuracy, the mean absolute error and relative error of total energy consumption were calculated using equations 8 and 9.

$$mean\ absolute\ error = \frac{\sum_{n=1}^{N} |\widehat{y_t} - y_t|}{N} \quad (8)$$

$$relative\ error\ in\ total\ energy = \frac{|\widehat{E} - E|}{\max(E, \widehat{E})} \quad (9)$$

where

y_t =appliance actual power at time t

$\widehat{y_t}$ = appliance estimated power at time t

E = total actual energy

$\widehat{E}$ = total predicted energy

3. RESULTS AND DISCUSSIONS

The three trained neural networks were evaluated for each appliance using the 10% of the data which was set aside for testing. To determine the error, the predicted and actual appliance power consumption measured from plug-level meter was compared for the oven, microwave and the air conditioner during one cycle of appliance activation as shown in Figure 3.

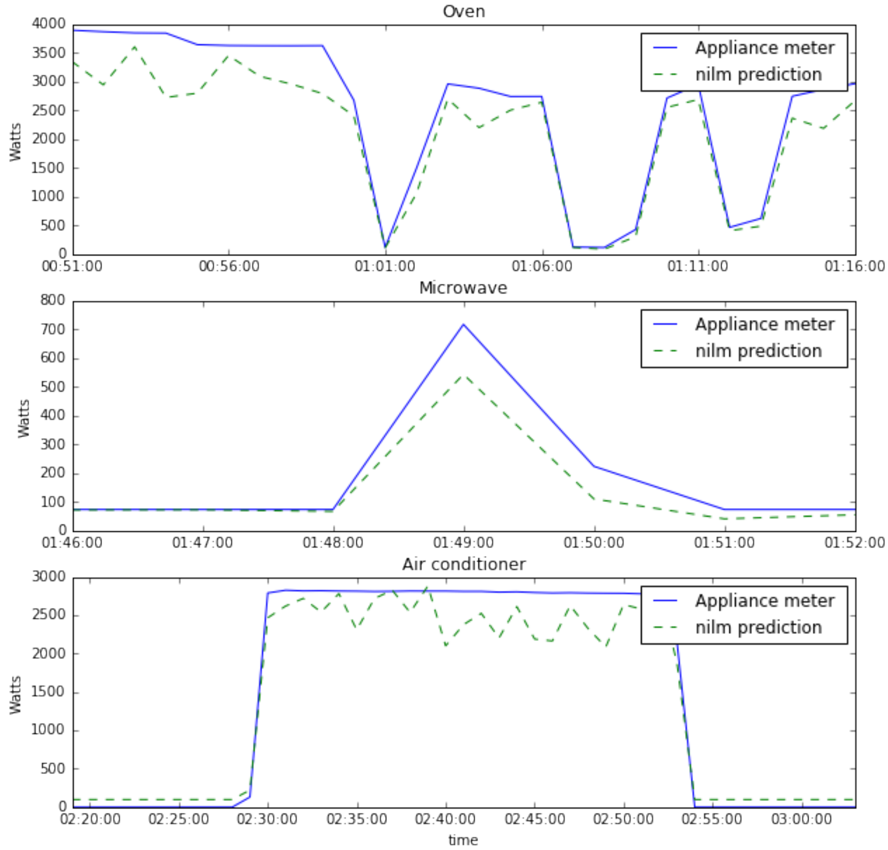

Figure 3. Disaggregation performance, comparing prediction by non-intrusive load monitoring and the mea rued power consumption (watts) at the plug-level. (a) Oven, (b) Microwave, (c) Air conditioner.

The disaggregation performance is summarized in Table 1 presenting the mean absolute error and relative error in total energy consumption for three aforementioned appliances calculated using equation 8 and 9. As it can be seen there is a good agreement between predicted and measured data and the accuracy is above the mean appliance energy.

Table 1.Dissagreagtion performance comparing the mean absolute error and relative error in total energy.

	Air conditioner	*Oven*	*Microwave*
Mean absolute error (Watts)	150.2	61.3	9.49
Relative error in total energy	32%	23%	14%

One the strengths of this study is using the Dataport Dataset which has aggregate and appliance level data for total of 669 buildings for duration of one month. Previous studies on non-intrusive load monitoring only included small dataset with limited number of buildings (Berges; Goldman et al., 2010; Kelly & Knottenbelt, 2015). It is worth mentioning that only 192 buildings had a meter for oven while microwave and the air conditioner data was available for 337 and 582 buildings, respectively.

The large Dataport dataset allowed us to train the convolutional neural nets on many samples which resulted in the models that generalized well. Each Convolutional networks has to train millions parameters depending on the size of the input image. In the case of air conditioner, the input shape of images were 100×100 pixels, and the convolutional network had 269M parameters, while the shape of activation signal for the microwave was only 30×30 pixels and the convolutional neural network had respectively 1.7M parameters. Although, one of the advantages of convolutional neural networks are shared parameters and consequently smaller number of parameters, the resulted models in this study were considerably large and training the models requires significant time and large storage for storing the models parameters.

4. CONCLUSIONS

In this study, we trained three convolutional neural networks for three appliances, an oven, microwave and air conditioner. The models generalized well on the unseen data which was set aside for testing. The predicted accuracy's for all three models were above mean energy prediction accuracy. This work was the initial step in using large datasets for disaggregation task. It will be helpful to compare the performance of the developed models to other established NILM algorithm such as Factorial Hidden Markov models or combinatorial optimization.

REFERENCES

Batra, N., J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh and M. Srivastava (2014). NILMTK: an open source toolkit for non-intrusive load monitoring. *Proceedings of the 5th international conference on Future energy systems.* Cambridge, United Kingdom, ACM**:** 265-276.

Berges, M. E., E. Goldman, H. S. Matthews and L. Soibelman (2010). "Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring." *Journal of Industrial Ecology* 14(5): 844-858.

Carrie Armel, K., A. Gupta, G. Shrimali and A. Albert (2013). "Is disaggregation the holy grail of energy efficiency? The case of electricity." *Energy Policy* 52: 213-234.

Darby, S. (2008). "Energy feedback in buildings: improving the infrastructure for demand reduction." *Building Research & Information* 36(5): 499-508.

Franco, J. (2008). "EIA: World Energy Consumption to Grow 50% by 2030." *Global Refining & Fuels Report* 12(14).

Goodfellow, I., A. Courville and Y. Bengio (2015). Deep Learning.

Kelly, J. and W. Knottenbelt (2015). Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. Seoul, South Korea, ACM**:** 55-64.

Lecun, Y., L. Bottou, Y. Bengio and P. Haffner (1998). "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86(11): 2278-2324.

Parson, O., G. Fisher, A. Hersey, N. Batra, J. Kelly, A. Singh, W. Knottenbelt and A. Rogers (2015). Dataport and NILMTK: a building data set designed for non-intrusive load monitoring. 1st International *Symposium on Signal Processing Applications in Smart Buildings at 3rd IEEE Global Conference on Signal & Information Processing.*

Sainath, T. N., A. r. Mohamed, B. Kingsbury and B. Ramabhadran (2013). Deep convolutional neural networks for LVCSR. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*.

Wang, Z. and T. Oates (2015). Imaging time-series to improve classification and imputation. *Proceedings of the 24th International Conference on Artificial Intelligence.* Buenos Aires, Argentina, AAAI Press**:** 3939-3945.