A Non-intrusive loading monitoring method based on Attention mechanism and denoising-Autoencoder

Qiuyu Duan¹,², Fei Zhang¹,², *, Xiaoying Ren¹,², Lu Gao¹

¹School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia Autonomous Region, China
²School of New Energy, North China Electric Power University, Beijing, China

*email: 2020022133@stu.imust.edu.cn

Abstract: This paper built denoising-Autoencoder (dAE), Attention-dAE and transformer-based regression models based on Non-Intrusive Load Monitoring (NILM) and Deep Learning (DL) methods, and uses them in UK-DALE proving on the dataset, the verification results show that the recognition accuracy of adding Attention mechanism and dAE is significantly improved, and the model parameters are greatly reduced.

1. Introduction

In the context of environmental pollution and energy revolution, clean and efficient electric energy plays a vital role in the energy supply mode. Various countries, regions, have launched in-depth discussions and research on the above issues. The governments and some organizations of various countries have set some deadlines for this issue. For example, China announced the target of 3060 carbon peak and carbon neutral. Relevant researchers have carried out research on all aspects of the power system. The use of new energy technologies in power generation can reduce carbon emissions. High voltage transmission technology is used to reduce line loss and improve transmission efficiency. The dynamic price policy is adopted on the user side to realize the high-efficiency energy consumption of users [2].

NILM refers to an algorithm, which can obtain accurate electricity consumption characteristics and habits by analyzing the electricity consumption data collected from the user's electricity meter without entering the user's home. Finally, it will guide the power companies to give more reasonable power supply policies to achieve safe, stable, efficient and low-carbon power supply.

NILM was proposed and studied by Hart in the 1980s. Its literature was published in 1992. It vectorized the active and reactive power waveforms and other electric quantities under different loads, and then mapped them into two-dimensional space. Finally, the power type is determined by comparing with the standard power phasor value. This method can be more accurately distinguish the types of electrical appliances with a large power difference, but for electrical appliances with a small power difference, the recognition accuracy cannot be guaranteed [3]. By the beginning of this century, there had been many studies combining traditional statistical methods with state quantity (time, temperature, etc.) to realizing Accurate identification. Besides, in some literature, the authors improve the sampling frequency to realize an Accurate collection of the power wave shapes, which are able to distinguish purely resistive electrical appliances (such as lighting equipment) and electrical equipment with inductive or capacitive loads (hair dryers or televisions, etc.). However, it can’t accurately distinguish
appliances with large different powers [3-6].

traditional statistical methods were the most effective before DL appeared in NILM, but the methods had higher pre-processing requirements and general performance [7]. Before and after 2010, some researches have carried out the application of DL in NILM, aiming at accurate recognition. In 2014, Dr. Kelly and others used RNN, dAE, LSTM and CO, FHMM to comparing, and selected six appliances in UK-DALE dataset [8]. The correctness of the identification is verified by combining the simulation bus data with the data of appliances at the same time. The results show that the three DL methods have good performance in seven evaluation indexes [9].

Based on the results of NILM and DL methods, this paper focuses on the attention mechanism of dAE, and selects dAE and the regression model based on transformer as the comparison to provide the basis for the horizontal comparison of the model performance.

2. Materials and Methods

2.1 Materials

2.1.1 UK-DALE
The low-frequency part of the UK-DALE dataset, contains the electricity meter data of 5 buildings. Among the data of each building, it includes the data of the total meter and the sub-meter data of the appliances that are not identical.

2.1.2 Data preprocessing
Selecting appliance. Through the observation and analysis of the electrical load data of the UK-DALE dataset, kettles, fridge, microwave ovens, and washing machines are selected as the datasets for the study. Seeing Table 1 for details of the number of specific start-stop events.

| appliances   | Building 1 | Building 2 | Building 3 | Building 4 | Building 5 |
|--------------|------------|------------|------------|------------|------------|
| kettles      | 2836       | 543        | 44         | 716        | 176        |
| fridge       | 16336      | 3526       | 0          | 4681       | 1488       |
| microwave    | 530        | 53         | 0          | 0          | 51         |
| washing machines | 3266   | 387        | 0          | 0          | 28         |

Extract electrical start-stop events. For the data processing during the start-stop events of the four electrical appliances, this article directly applies the modules and functions in NILMTK to all aspects of data preprocessing, such as start-stop events time detection, slicing, etc. [10-11].

Production of pytorch dataset.

1) Divided the total meter data according to activation, and divided the total meter data into two types of total meter data with and without target appliances.

2) Sequence length selection. Taking the kettle as an example, the step length of the load waveform is maintained at about 120. To avoid more complicated calculations caused by inconsistent compensation, the activation data is supplemented to a step length of 160, and its position is not in the new data. Fixed, determined randomly.

3) 50% random is taken with activation, and the other 50% is taken without activation. The ground-truth with activation is the output of the appliance at the corresponding time, and the ground-truth without activation is 0. The advantage of this processing is that the generated identification model can not only accurately identify the bus data with start-stop events, but also correctly identify the electrical equipment without start-stop events or no electrical load.

2.1.3 Setting of training set and test set
The training set and test set data selected in this paper are shown in Table 2.
Table 2. Details of the allocation of training set and test set

| appliances   | Building1 | Building2 |
|--------------|-----------|-----------|
| kettles      | 1, 2, 3, 4| 5         |
| fridge       | 1, 2, 4   | 5         |
| microwave    | 1, 5      | 2         |
| washing machines | 1, 2     | 5         |

2.2 Methods

2.2.1 dAE

Typical denoising tasks include removing grain from old photos, removing reverberation from audio recordings, and even filling in masked parts of the image. The processing of NILM is essential to recover the power load of the target electrical appliance from the summary meter data containing irrelevant loads of other electrical appliances. Neural networks commonly used for "denoising" tasks are called dAE [4].

![Figure 1. Neural network structure](image)

The neural network of the dAE built in this paper, starts with 2 convolutional layers, and the purpose of convolution calculations is to extract data features. The more convolutional layers, the more features are extracted. However, during the debugging process, it was found that when the convolutional layer is set to 2 layers, the best feature extraction and relatively reasonable calculation speed can be obtained. In the second step, the features extracted by the convolutional layer are transferred to the 3 fully connected layers. Finally, after two convolutional layers, the iterative result is output. As a kind of encoder-decoder model, dAE also has a completely symmetrical encoding and decoding process [12].

2.2.2 Attention-dAE

Attention mechanism. Attention mechanism is to map query and key to the same high-dimensional space to calculate similarity, while multi-head attention maps query and key to different subspaces of the high-dimensional space to calculate similarity.
The Multi-head Attention mechanism is to map the same query, key, and value to different subspaces of the original high-dimensional space to calculate the attention under the condition that the total amount of parameters remains unchanged. This reduces the dimensionality of each vector when calculating Attention mechanism of each head, and prevents overfitting in a sense. Since Attention mechanism has different distributions in different subspaces, Multi-head Attention actually finds the correlation between different angles of sequences. And in the final concat step, the association relationships captured in different subspaces are combined again.

Attention-dAE. Combining Attention mechanism (this article is a multi-head attention mechanism) with dAE, the network structure is shown in Figure 2, where the number of heads of Attention mechanism is set to 3.

2.2.3 Transformer
Transformer is a DL built based on Attention, which can not only achieve the same function as RNN, has a better effect. It has become the most popular neural network in recent years. [13]

The network structure of the transformer built in this article is shown in Figure 3. After debugging the multi-head attention mechanism here, the number of attention heads is 16.
3. Results & Discussion

3.1 Results
After completing the calculation of the above four electrical appliances and a total of twelve load identification models, this article would compare the model size (the feasibility of integrating into a smart meter), the Mean Absolute Error (MAE), the Mean Square Error (MSE), and the Relative Load of the total load. The performance evaluation of the NILM methods composed of several indicators such as error and correct recognition rate and each neural network.

The Mean Absolute Error (MAE) reflects the average value of the absolute value of the error between the observed value and the true value. When the predicted value is completely consistent with the true value, it is equal to 0, which is a perfect model. The greater the error, the greater the value.
\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \]  
(1)

Where \( \hat{y}_i \) represents the i-th true value. \( y_i \) represents the i-th predicted value.

Mean Square Error (Mean Square Error): when the predicted value is completely consistent with the true value, it is equal to 0, that is, a perfect model. The greater the error, the greater the value.

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \]  
(2)

The meaning of each element in the formula is the same as the above formula.

Relative Error in Total Energy:

\[ REITE = \frac{|E - \hat{E}|}{\max(E, \hat{E})} \]  
(3)

Proportion of total energy correctly assigned.

\[ POTECA = 1 - \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} |\hat{y}_t^{(i)} - y_t^{(i)}|}{2 \sum_{t=1}^{T} \hat{y}_t} \]  
(4)

3.1.1 Model size comparison

The specific statistical results are shown in Table 3.

| appliances   | dAE     | Attention-dAE | transformer |
|--------------|---------|---------------|-------------|
| kettles      | 1380721 | 724125        | 11413136    |
| fridge       | 4269937 | 2168733       | 48862064    |
| microwave    | 2037361 | 1052445       | 18357536    |
| washing machines | 12379441 | 6223485      | 249296636   |

As the model size can be seen from the table 3, the model size of Attention-dAE is the smallest among the three DL, accounting for about half the size of the dAE model on average, one-twentieth of the Transformer. Meanwhile, as the input data step length becomes longer, the parameters of models and the requirements for the number of GPU also increasing.

3.1.2 Comparison of evaluation indicators

This summary will explain and analyze the comprehensive performance evaluation results of the NILM model generated by each neural network.
The following conclusions can be drawn from the differences in the indicators in Figure 4:

1) Among the error indicators, Attention-dAE and Transformer have better identification performance in most cases, but some models are not ideally realized, which may be caused by factors such as insufficient data sets.

2) In the identification accuracy, the performance of Attention-dAE among the three models is better than that of dAE, while the performance of Transformer is 3% higher than that of Attention-dAE.

3.1.3 Identification result

The effect can be seen in Figure 5. Among these, the load with larger power and start-stop characteristics obviously can have close real data, such as the kettle. The load with longer and fluctuations frequently have less predicted ideal results, such as the washing machine and microwave. But these models are still able to predict the start-stop actions of electrical equipment accurately.

3.2 Discussion

Through comparison, we found that the comprehensive performance of Attention-dAE was the best (model size, accuracy, etc.). Thus, Attention-dAE is a more effective NILM method at present. However, there are still insufficient conditions that can lead to the performance degradation of Attention-dAE.
4. Conclusions
1) The research of this paper is only for the bus data collection. In the future, the correlation between other electrical quantities and non-electrical quantities (such as temperature and time) can be expanded.
2) With the continuous upgrading of hardware, Transformer with a large number of model parameters will inevitably become a stable NILM method popularized to millions of households.
3) In the future, relevant research can be carried out for more sufficient data sets to avoid poor performance due to insufficient data.

References
[1]Yiming Liu, Huimin Li, Leting wang & Hasan RAFIG. Non-invasive load monitoring method based
on convolutional neural network. Electrical measurement and instrumentation.

[2] Hart, G. W. (1992). Nonintrusive appliance load monitoring. Proceedings of the IEEE, 80(12), 1870-1891.

[3] Kim, H., Marwah, M., Arlitt, M., Lyon, G., & Han, J. (2011, April). Unsupervised disaggregation of low frequency power measurements. In Proceedings of the 2011 SIAM international conference on data mining (pp. 747-758). Society for Industrial and Applied Mathematics.

[4] Berges, M. E., Goldman, E., Matthews, H. S., & Soibelman, L. (2010). Enhancing electricity audits in residential buildings with nonintrusive load monitoring. Journal of industrial ecology, 14(5), 844-858.

[5] Berges, M., Goldman, E., Matthews, H. S., & Soibelman, L. (2009). Learning systems for electric consumption of buildings. In Computing in Civil Engineering (2009) (pp. 1-10). Chang, H. H. (2012). Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses. Energies, 5(11), 4569-4589. Chang, H. H., Chen, K. L., Tsai, Y. P., & Lee, W. J. (2011). A new measurement method for power signatures of nonintrusive demand monitoring and load identification. IEEE Transactions on Industry Applications, 48(2), 764-771.

[7] Siyun Chen, Feng Gao, Ting Liu, Qiaozhu Zhai, & Xiaohong Guan. (2016). Loading decomposition method and sensitivity analysis based on factorial hidden Markov model. Automation of Electric Power Systems, 40(21), 128-136.

[8] Kelly, J., & Knottenbelt, W. (2015). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. Scientific data, 2(1), 1-14.

[9] Kelly, J., & Knottenbelt, W. (2015, November). 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 (pp. 55-64).

[10] Kelly, J., Batra, N., Parson, O., Dutta, H., Knottenbelt, W., Rogers, A., ... & Srivastava, M. (2014, November). Nilmkt v0. 2: a non-intrusive load monitoring toolkit for large scale data sets: demo abstract. In Proceedings of the 1st ACM Conference on Embedded Systems for Energy-efficient Buildings (pp. 182-183).

[11] Batra, N., Kelly, J., Parson, O., Dutta, H., Knottenbelt, W., Rogers, A., ... & Srivastava, M. (2014, June). NILMTK: An open source toolkit for non-intrusive load monitoring.

[12] Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008, July). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine learning (pp. 1096-1103).

[13] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.