Intelligent Control and System Engineering 2024, 2(1), 427.
https://doi.org/10.59400/icse.v2i1.427

Article

End-to-end NILM model of industrial power data based on autoencoder transformer

Ce Li[1,2], Fanglin Guo[1], Rong Yang[1,*], Hua Wang[1], Bo Yang[1]

[1] Energy Internet Division, State Grid Gansu Electric Power Company, Lanzhou 730000, Gansu Province, China
[2] School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, Gansu Province, China
*** Corresponding author:** Rong Yang, yyyyinny99@gmail.com

CITATION

Li C, Guo F, Yang R, et al. End-to-end NILM model of industrial power data based on autoencoder transformer. Intelligent Control and System Engineering. 2024; 2(1): 427. https://doi.org/10.59400/icse.v2i1.427

ARTICLE INFO

Received: 22 December 2023
Accepted: 17 January 2024
Available online: 28 February 2024

Abstract: Energy detection is an important part of intelligent power consumption, and its key technology is non-intrusive load monitoring (NILM). In this study, an end-to-end model is proposed to realize the NILM of commercial power data using the autoencoder-based transformer method. Firstly, we measured the operating power of different electrical appliances across different modes and combined the operating modes of electrical appliances. Considering the relatively large number of industrial electrical appliances, to ensure accuracy, we used Autoencoder to recode and reduce the dimension of the combined coding. Secondly, the transformer model was used to train the translation of the total power consumption information sequence and the state sequence of electrical appliances. Through our model, the electrical signals to be decomposed can be translated into different electrical state codes so as to realize load energy decomposition. Finally, when our model was applied to the gas station field data, the accuracy was as high as 90.17%.

Keywords: non-intrusive load monitoring; transformer; deep learning; smart grid

1. Introduction

Climate warming is a global problem, and as the biggest environmental challenge in the 21st century, it has been widely regarded by society as one of the most important factors that may lead to human extinction [1]. A large amount of greenhouse gases produced by the thermal power generation of power plants is discharged into the atmosphere, which is one of the main reasons for climate warming. The power generation industry still accounts for the largest share of global greenhouse gas emissions because the main fuel for power plants in the process of thermal power generation is still fossil fuels. According to the latest research report on temperature and electricity [2], the total electricity consumption of society will increase with advanced ambient temperatures. Therefore, excessive energy consumption will trigger a series of chain reactions affecting the ecological environment. It is necessary to promote and carry out energy-saving work in an orderly manner to reduce energy consumption and improve energy utilization.

With the development of related technologies [3], the smart grid can effectively achieve energy savings [4], and building an efficient and available load monitoring system can help promote the rapid development of smart grids. As an indispensable part of the smart grid, the load monitoring system [5] has been developed by more scholars in recent years, and it is mainly aimed at the electricity load of the family house. The purpose of load monitoring is to obtain power information such as power consumption, current, and voltage of the residential load and its changes through smart meters so as to analyze the operation of the load and the power consumption [6]. This information can help power grid companies design effective and feasible algorithms to allocate resources [7], help users understand their own electricity consumption to improve their electricity consumption behavior [8], and identify malicious loads used by residents and prevent fires in advance [9].

Nonintrusive load monitoring (NILM) decomposes the power consumption data collected by the electricity meter, displays the usage of electrical appliances in real-time, and provides technical support for the smart grid [10]. In this paper, we propose a method that combines the autoencoder and the transformer, which effectively solves the problem of the large number and complex states of industrial appliances while ensuring high performance.

2. Related work

The non-intrusive monitoring technology is an energy decomposition method proposed by Hurt in 1992. The non-intrusive monitoring method is used to collect data and decompose it to obtain the energy usage of the user's electrical appliances [11]. After years of research, scholars have adopted different technologies based on the sampling frequency of non-intrusive monitoring data, mainly involving signal processing and machine learning. Previous studies [12–22] have conducted high-frequency sampling of power consumption data and used signal processing technology to propose analysis methods such as transient event detection, template matching filtering, and sliding window, respectively, which achieve high energy decomposition accuracy, but it is not suitable for power grids. High requirements for data acquisition, transmission, and storage are put forward, and in practice have great limitations. In contrast, the decomposition of the collected low-frequency data is more conducive to the application of NILM technology in smart grids. Low-frequency data does not have complete signal characteristics, and scholars have studied this topic through machine learning methods.

The algorithms that have been involved include support vector machine (SVM [16], K-nearest neighbor (KNN) [17]) artificial neural network (ANN) [18], Hidden Markov models [19,20] and other methods. The above machine learning methods have made some progress in the decomposition of non-intrusive low-frequency data, but the overall accuracy of the algorithm needs to be improved. In order to improve the accuracy, Batra et al. implemented the combined the factorial hidden Markov model algorithm (factorial hidden Markov models, FHMM) and optimization (CO) method to improve the accuracy of non-invasive monitoring, and made a toolkit NILMTK as a comparison algorithm for non-invasive monitoring accuracy [21,22], but the computational efficiency of the combined model is low, and with an advanced number of household appliances, more complex models are difficult to train. Recently, Wittmann proposed mixed-integer linear programming (MILP) for NILM, which achieved high accuracy [22].

The above research methods did not consider the time-related characteristics of electrical signals. At the same time, with recent developments in deep learning technology, the transformer model proposed by Google has shined in both Natural Language Processing (NLP) and Computer Vision (CV) fields due to its good performance. Considering that the electrical signal data itself is time-series data, in order to sufficiently utilize the time-series characteristics of electrical signals, a transformer method is proposed and introduced into our research. The whole network structure of Transformer is completely composed of the attention mechanism, which consists of self-attention and Feed Forward Neural Network [23]. It enables the network to better memorize the correlation information between the data before and after, thereby providing time scale features for load decomposition.

The sequence translation realizes a one-to-one mapping, and the energy is decomposed into the mapping of a total signal to multiple electrical signals. In order to realize the application of the deep sequence translation model, firstly, the combined coding of each mode of the electrical appliance is performed, and all electrical appliances are represented by a state code, so as to realize the dimension reduction representation of the electrical appliance state. Different from previous studies, the research application scenario of this paper is industrial electricity rather than household electricity. The characteristics of industrial electricity include a large amount of electrical data and complex electrical status. In order to further ensure the accuracy of the method, at the same time, we introduce self-encoding to re-encode and learn the electrical state encoding. Finally, this paper uses real data from a gas station in a city in Gansu Province to verify the proposed method, and compares it with the existing algorithms, our proposed method has higher accuracy.

3. Dataset acquisition and preprocessing

3.1. Dataset acquisition

Because there are few public data sets of non-invasive load decomposition in the industrial scene, we collected the power consumption data for each appliance in a gas station through a set of sensor devices. Considering that the Internet of things equipment cannot be installed directly in the operation area of the gas station, it is selected to be installed on the switch in the general distribution box close to the transmission end of the distribution room. It is different from the traditional non-invasive load decomposition data acquisition method of installing acquisition equipment at the end of electrical equipment. In a realistic gas station scenario, where many appliances are on an assembly line and switches in the distribution box can control one or more electrical appliances depending on the business situation, our acquisition solution is realistic.

We collected data on the 12 appliances in the gas station, these were canopy lights, freezer, canopy lights strip, oil-submerged pump, convenience store socket, central air-conditioning, kitchen socket, counter socket, integrated office socket, gas station outdoor advertising signage + inlet and outlet lightbox (OAS + IOLB), uninterrupted power supply (UPS), and lounge socket.

The acquisition equipment uses a Snap-on current transformer, which is directly clamped on the power wire of the equipment. The acquisition equipment is shown in **Figure 1**. We collected data on the power consumption of each appliance at two gas stations in the same area, and each equipment generates a power consumption record every 15 min. The collection time is 8 months, and a total of 324,000 data are collected. The acquisition indicators include electric quantity, current, power factor, reactive power, active power, etc. In this paper, we use active power as the power state index of consumers.

Figure 1. Data acquisition equipment.

We obtained the total power consumption data from the data center of the power company of Gansu Province. In order to detect the reliability and consistency of data acquisition, **Figure 2** compares the data collected by the sensor with the data of the electricity meter. The trend and amplitude of the total power are consistent, which proves that the data collected by the sensor is effective.

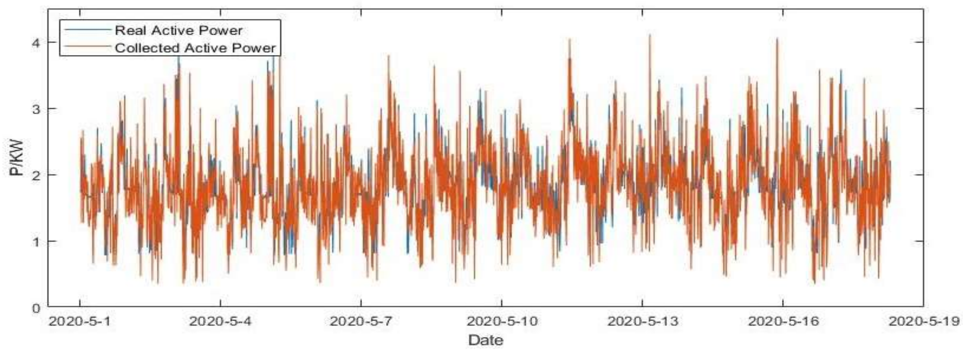

Figure 2. Comparison of collected power data and real power data.

3.2. States code of electric device

The working mode of electrical appliances used in gas stations is relatively stable. For example, the working state of electric submersible centrifugal pump can be basically divided into three working modes, and the electrical appliances have rated power under each working mode. Although there will be some fluctuations in the actual working process, through experimental analysis, under each working mode, the appliance can approximately obey Gaussian distribution, and the distribution variance is small enough.

In order to obtain the power distribution of each working mode in the actual use of electrical appliances, this paper obtains the power value distribution of each electrical appliance by creating probability mass function (PMF) on the sampling data. The power distribution and electrical modes are divided correspondingly, and the mean value of power distribution under each electrical working mode is taken as the representation of the actual power under its working mode. The actual power of each appliance is constructed into an index table, as shown in **Table 1**. **Figure 3** shows the PMF diagram of electrical appliances in different states.

Table 1. Index diagram of electric device states and actual power.

Device	Number of states	A	B	C
2ABC	3	0.17	1.85	3.28
3ABC	2	0.04	3.73	-
4C	2	0	1.11	-
5A	1	0.67	-	-
5B	1	0.67	-	-
5C	1	0.67	-	-
6A	2	0	1.04	-
6B	2	0	1.04	-
7A	3	0.23	0.91	1.54
7B	3	0.23	0.91	1.54
7C	3	0.23	0.91	1.54
8A4B8C	2	0	0.61	-

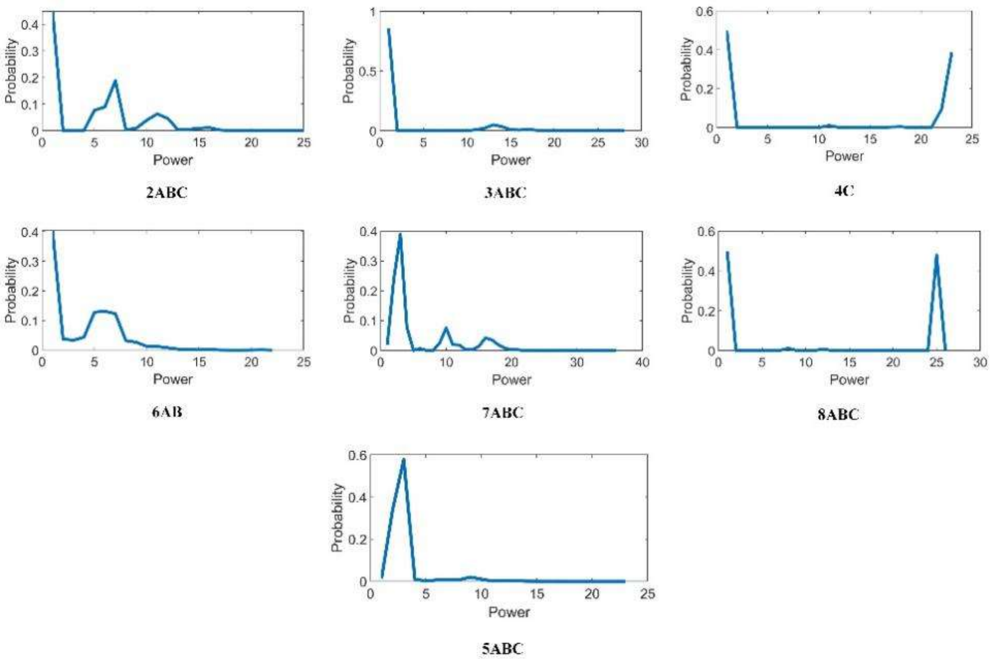

Figure 3. The PMF of different device.

The non-invasive monitoring method generates a decomposition signal through the total energy to restore the electrical energy signal. The energy consumption of each equipment is obtained by accumulating each electrical signal. For the time t, the expression of non-invasive monitoring is shown in Equation (1).

$$f(X_t) = [Y_t^1, Y_t^2, \dots, Y_t^M] \tag{1}$$

where X_t is the total power consumption at time t, and Y_t^i is the power value of the i-th electrical appliance decomposed at time t. The M is the number of electrical appliances, and f is the NILM method.

Electrical loads are divided into four basic categories: switching, continuous activity, continuous variable, and limited state [24]. Due to the limited operating states of the appliance and the rated power in each operating state, Equation (1) can be transformed into Equation (2):

$$f(X_t) = [S_t^1, S_t^2, \dots, S_t^M] \tag{2}$$

where S_t^i means the working states of the *i*-th appliance at time *t*.

We can further define the working state code *y* of *M* appliances at some point as Equation (3):

$$y = [S^1, S^2, \dots, S^M], S^i \in \{0,1,\dots,n_i - 1\} \tag{3}$$

where S^i is the working state of the *i*-th appliance and n_i represents the number of working state of the *i*-th appliance. The total number of working state code is $\prod_{i=1}^{M} n_i$.

Finally, the NILM problem could be transformed into Equation (4).

$$y = f(X), y \in \{0,1,\dots,N - 1\} \tag{4}$$

where *X* is the total power consumption data of the appliances, *f* is the NILM method, *y* is the state code and *N* is the working state code number.

4. Method

Our model composed of three parts: a states encoder, a total power consumption encoder and a state decoder. The state encoder is composed of a double-layer autoencoder, which is responsible for transforming the dense vector representing the state of electrical appliances into sparse vector.

Total power consumption encoder and state decoder respectively represent the encoder and decoder modules in the transformer network, which is composed of *N* electric quantity encoders and state decoders. By encoding the total electric quantity and decoding the reduced dimension state vector, the process of non-invasive load monitoring is realized.

4.1. Code the combined states code with autoencoder

Previous tasks mainly focused on the scene of household appliances with a small number of appliances. In this paper, the gas station data we collected contains the power consumption status of 12 consumers in total. However, the total available power consumption data is only three data: total power consumption, total power, and total voltage. In order to solve the problem that the accuracy of the model is too low due to learning the law of dense encoded data from sparse encoded data, we introduce the Autoencoder to reduce the dimension of the state code of electrical appliances. The network topology of the autoencoder is shown in **Figure 4**.

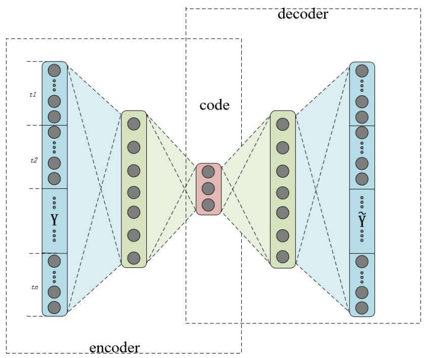

Figure 4. Structure of autoencoder.

The autoencoder (AE) is a neural network that uses the back propagation technique to make the output identical to the input value. It compresses the data into a latent space representation before reconstructing the output using this representation.

Autoencoder consists of encoder and decoder. The encoder can compress the input into a latent space representation, which can be expressed as Equation (5):

$$h = f(x) \tag{5}$$

The decoder reconstructs the input from a possible spatial representation in part, which can be represented as Equation (6):

$$\gamma = g(h) \tag{6}$$

Therefore, the whole autoencoder can be described as Equation (7):

$$g\big(f(x)\big) = r \tag{7}$$

where r is the same as x, and x is the original input.

Copying the input to the output appears to be pointless, but usually we don't care about the output of the decoder. On the contrary, what we want is to make h obtain useful features by training the autoencoder to copy the input. The autoencoder whose coding dimension is smaller than the input dimension is called incomplete autoencoder if one option to extract usable features from the autoencoder is to limit the size of h to be smaller than that of x. Learning the incomplete representation will force the autoencoder to capture the most significant features in the training data. The learning process is to minimize the reconstruction error (Equation (8)):

$$L\left(x, g\big(f(z)\big)\right) \tag{8}$$

where L is a loss function used to calculate the mean square error between x and $g\big(f(z)\big)$.

Automatic coder is generally used for data dimensionality reduction or feature learning, which is similar to PCA, but automatic coder is much more flexible than PCA, because it can represent both linear transformation and nonlinear transformation.

Therefore, by training the autoencoder, we can reduce the dimension and copy the state combination data of consumers. In this paper, we take the power consumption data of a day as a batch and collect the data every 15 min, so the data of a batch is 96-time nodes. As shown in **Figure 4**, the combined code of the state of the consumer for one day of input data. Through the autoencoder, we obtain a new 6-dimensional code, and the code dimension is consistent with the input data dimension.

4.2. Transformer for NILM

This section provides background information on transformer, which serves as the foundation for our model. The scaled dot-product attention is the foundation of transformer. Therefore, we'll start there. Given a query $q_i \in \mathbb{R}^d$ from all T'queries, a set of keys $k_t \in \mathbb{R}^d$ and values $v_t \in \mathbb{R}^d$ where t = 1, 2, ..., T, the weighted sum of values v_t produced by the scaled dot-product attention is decided by the dot-products of query q and keys k_t. In practice, we use the matrices $K = (k_1, ..., k_T)$ and $V = (v_1, ..., v_T)$ to record k_t and v_t. Equation (9) shows the attention output for query q:

$$A(q_i, K, V) = V \frac{\exp\left(k^T q_i \,/\, \sqrt{d}\right)}{\sum_{t=1}^{T} \exp\left(k_t^T q_i \,/\, \sqrt{d}\right)} \tag{9}$$

The multi-head attention is made up of H paralleled scaled dot-product attention layers termed "head", each of which is a separate dot-product attention layer. The following is the outcome of multi-head attention (Equations (10) and (11)):

$$MA(q_i, K, V) = W^O \begin{pmatrix} \text{head}_1 \\ \ldots \\ \text{head}_H \end{pmatrix} \tag{10}$$

$$\text{Head}_j = A(W_j^q q_i, W_j^K K, W_j^V V) \tag{11}$$

where $W_j^q, W_j^K, W_j^V \in \mathbb{R}^{\frac{d}{H} \times d}$ are the independent head projection matrices, j = 1, 2, ..., H, and $W^O \in Rd \times d$.

This is a very wide expression of attention. for example, when the query is the decoder's hidden states, and both the keys and values are all the encoder's hidden states, it reflects common cross-module attention. Another example of multi-head attention is self-attention, in which the queries, keys, and values all come from the same hidden layer (see also in **Figure 5**).

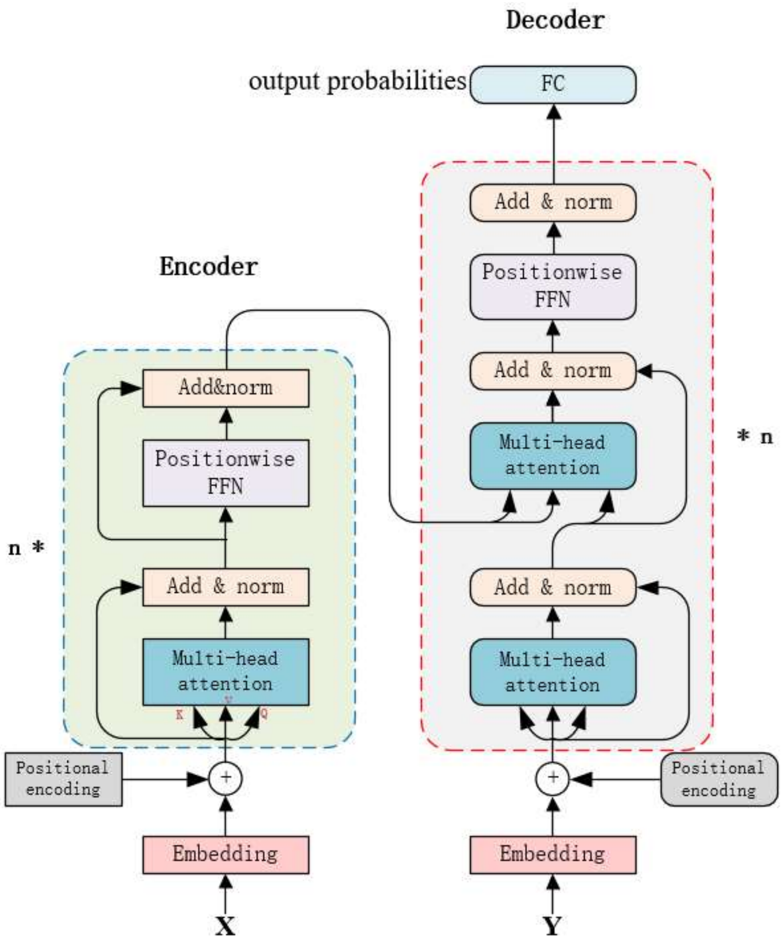

Figure 5. Transformer with n-layer encoder and n-layer decoder.

Transformer network is a machine translation-based encoder-decoder model. Multi-head attention and a pointwise feed-forward layer are the foundations of transformer [23]. The input from the multi-head attention layer is sent via two linear projections with ReLU activation before being passed through the pointwise feed-forward layer. Two convolution layers with kernel size one can be seen as the feed-forward layer. Transformer's encoder and decoder are made up of numerous such building units with the same number of layers. Each layer's decoder receives input from its own encoder as well as the output of the lower layer decoder. Both the encoder and the decoder use self-attention. Cross-module attention is also used between the encoder and the decoder. To keep the auto-regressive property, the decoder's self-attention layer can only pay attention to the current and prior positions. All input and output layers have a residual connection [25]. All layers are also subjected to normalization layer [26] (NormLayer). One layered transformer is used in this article.

5. Experiments and result

The input data is the total power consumption data of the gas station provided by the State Grid of Gansu Province, in which the data quality of total power consumption, total power, and total voltage meet the requirements of the laboratory. The input data set is formed by noise reduction processing: $X = \{Q, P, V, T\}$, $Q = (q_1, q_2, \ldots, q_1)$, $P = (p_1, p_2, \ldots, p_1)$, $V = (v_1, v_2, \ldots, v_1)$, $T = (t_1, t_2, \ldots, t_1)$. Where Q, P, V and T represent electric quantity, power, voltage, and acquisition time respectively, At the same time, our data are collected every 15 min, so $i = 96$ in the data every day.

The sequence dimension of electrical appliance status label is $n * i$ dimension. Where $n = 12$ is the number of electrical devices, $i = 96$ in the data every day. The neurons number in the two layers of self encoder is 10 and 6 respectively, with dropout rate is 0.1, learning rate is 10^{-4}, and Adam for optimization. The dimension of the power on state sequence is reduced from the encoder to 6 dimensions, and then used as the input of the state decoder in the transformer. The data set contains 196 days of data, of which 80% is used as training set and 20% is used as test set.

As in previous studies, we used accuracy and mean absolute error (MAE) to evaluate the model [27–29]. Several state-of-the-art architectures were used to compare with our model, including BiLSTM and BiGRU [28] as well as a seq2seq model [30], which were modified to have the same input length and maximum hidden size as our model and used the same training strategy. **Table 2** shows the results of the comparison on our dataset. We compared the predicted output of the refrigerator with the sampling graph of the real label, where the blue represents the actual power of the refrigerator at different times, and the orange represents the predicted result. As shown in **Figure 6**, our model has a better fitting effect compared to other commonly used deep learning models.

Table 2. Model performances on different device with different mothed.

Device	Model	ACC	MAE
Submersible pump	BiGRU	0.78	40.2
	BiLSTM	0.78	41.2
	Seq2seq	0.79	35.2
	Ours	0.86	30.5
UPS power supply	BiGRU	0.92	27.6
	BiLSTM	0.98	35.4
	Seq2seq	0.97	36.4
	Ours	0.99	34.9
Light belt	BiGRU	0.97	21.8
	BiLSTM	0.97	17.5
	Seq2seq	0.98	18.5
	Ours	0.98	18.5

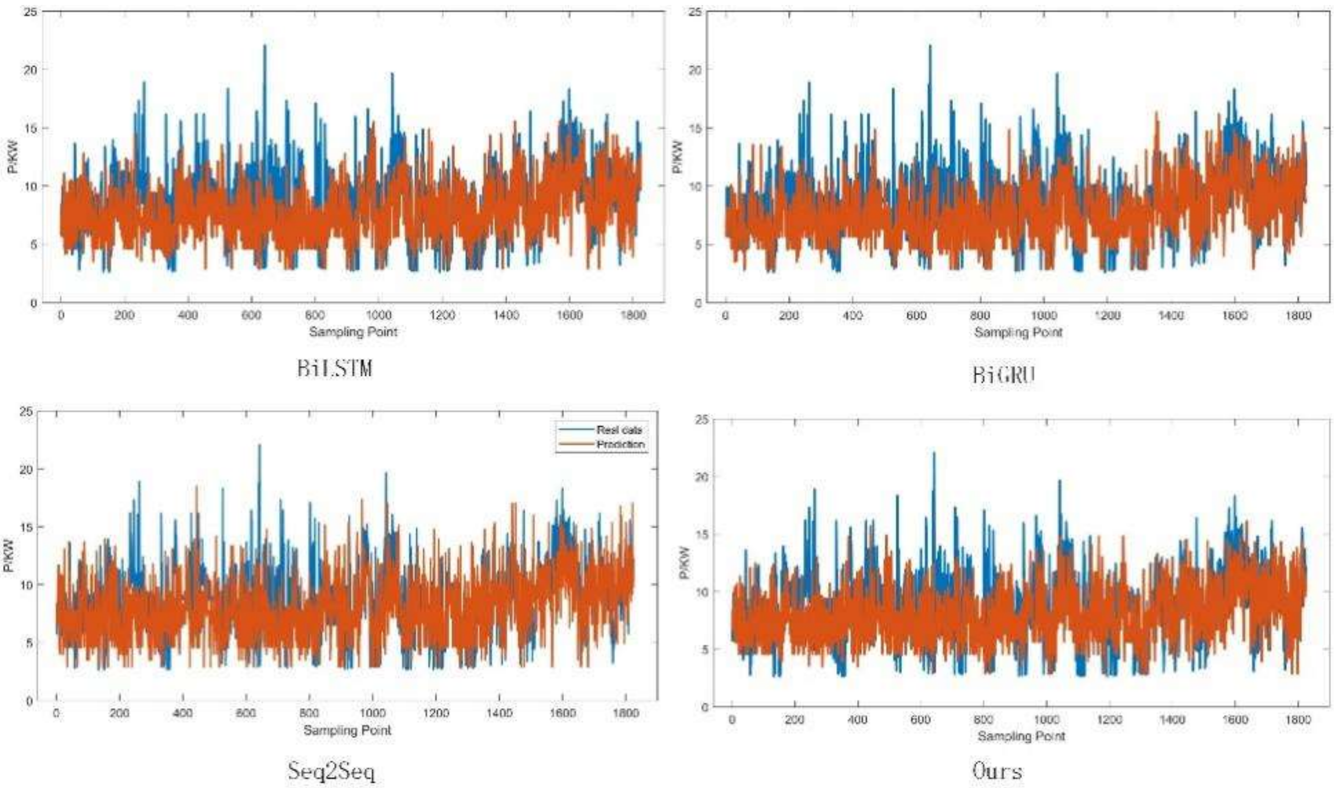

Figure 6. Sample output of refrigerator.

Figure 7 compares the different models with a sample of the refrigerator and the graph shows that our model is more accurate as well as stable.

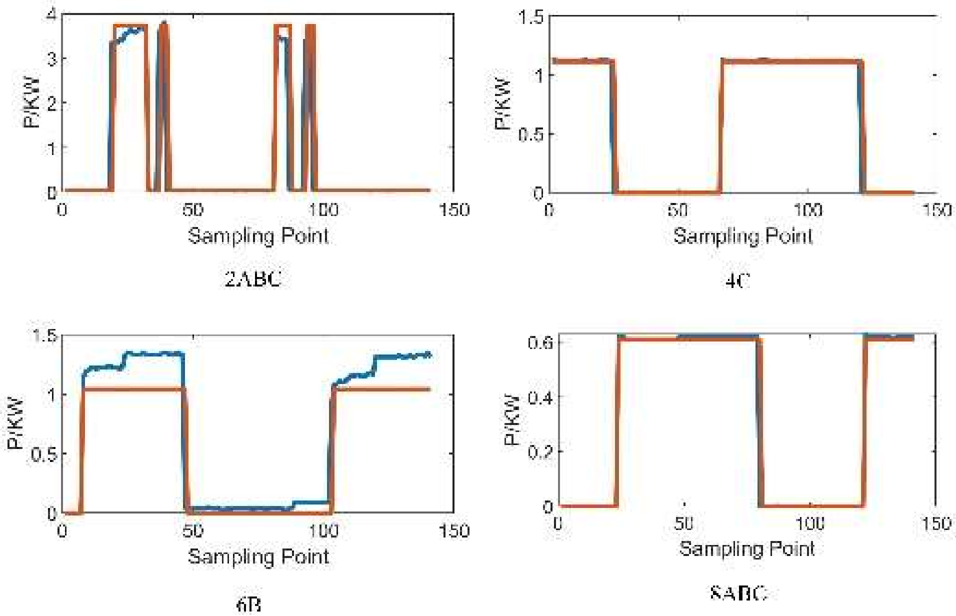

Figure 7. Real and predicted power distribution of different devices.

The results show that the average performance of our model is better than other models and is superior on most devices.

Our model is an end-to-end model and has good performance. As shown in **Table 3**, NILM accuracy of our method on industrial electrical equipment reaches 90.17%. Among them, the accuracy of equipment 6AB: Lounge Socket & OAS + IOLB is relatively low, mainly due to the irregular electricity consumption law of these equipment and inconvenient to collect.

Table 3. Model performances on different appliance.

Device	Device	Accuracy
2A2B2C	Oil-submerged pump	0.86
3A3B3C	Central air-conditioning	0.89
4C	Canopy lights strip	0.96
5B	Kitchen socket	0.99
5C	Integrated office socket	0.99
6B	OAS + IOLB	0.76
5A	UPS	0.99
6A	Lounge socket	0.76
7A	Counter socket	0.88
7B	Convenience store socket	0.88
7C	Freezer	0.88
8A4B8C	Canopy lights	0.98
Total accuracy		0.901666667

In **Figure 7**, we present the distribution of real power consumption data and predicted data of four devices with different accuracy rates. Where the orange and blue lines represent the predicted power distribution, and the actual power distribution respectively.

As we suspected, 4C and 8ABC electricity consumption laws are relatively standard, so the prediction accuracy is higher. Meanwhile, the prediction accuracy of 2ABC and 6B is relatively low.

6. Conclusion

In this paper, a method of applying the sequence translation model composed of an autoencoder and transformer to the non-invasive load decomposition problem is proposed, which makes full use of the correlation between the amplitude characteristics of the power signal and the operation mode on the time scale to achieve the load energy decomposition. Verified by real industrial data, the proposed method achieves high accuracy.

a) For the low-frequency monitoring data collected, the NILM method to construct deep sequence translation can make full use of the correlation between the amplitude characteristics of the data and the operation mode on the time scale.
b) By combining the operation modes of electrical appliances into state codes, a reduction in sequence mapping is realized, which eliminates the redundancy of the load decomposition model and reduces the time of training. At the same time, the dimension reduction of the autoencoder can effectively solve the problem of the high dimension of multi-device state combination code.
c) The depth sequence translation model proposed in this paper has good scalability. When the number of electrical devices to be decomposed increases, the proposed method can still decompose the load data.

We were able to properly adapt the architecture for energy disaggregation; therefore, the autoencoder and transformer model is effective for NILM tasks. Future research could concentrate on developing a lightweight autoencoder-based transformer model to speed up training and inference, as well as a more efficient optimization procedure to enhance prediction quality on multi-staged appliances and unbalanced datasets.

Author contributions: Conceptualization, CL and FG; methodology, CL; software, CL and RY; validation, FG, RY and BY; formal analysis, CL and RY; investigation, CL; resources, CL; data curation, FG and RY; writing—original draft preparation, CL and RY; writing—review and editing, CL, FG and RY; visualization, HW and BY; supervision, HW; project administration, CL. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

1. Schelling TC. Some economics of global warming. The American Economic Review. 1992, 82(1): 1-14.
2. Santamouris M, Cartalis C, Synnefa A, et al. On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—A review. Energy and Buildings. 2015, 98: 119-124. doi: 10.1016/j.enbuild.2014.09.052
3. Jiang X, Ding H, Shi H, et al. Novel QoS optimization paradigm for IoT systems with fuzzy logic and visual information mining integration. Neural Computing and Applications. 2019, 32(21): 16427-16443. doi: 10.1007/s00521-019-04020-3
4. Tuballa ML, Abundo ML. A review of the development of Smart Grid technologies. Renewable and Sustainable Energy Reviews. 2016, 59: 710-725. doi: 10.1016/j.rser.2016.01.011

5. Zoha A, Gluhak A, Imran M, et al. Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. Sensors. 2012, 12(12): 16838-16866. doi: 10.3390/s121216838
6. Tsai MS, Lin YH. Modern development of an Adaptive Non-Intrusive Appliance Load Monitoring system in electricity energy conservation. Applied Energy. 2012, 96: 55-73. doi: 10.1016/j.apenergy.2011.11.027
7. Huaiying S, Yan L, Qi Z, et al. Research on the quality Data Service Platform architecture of Intelligent electricity Meters. Journal of Physics: Conference Series. 2020, 1693(1): 012120. doi: 10.1088/1742-6596/1693/1/012120
8. Masera M, Bompard EF, Profumo F, et al. Smart (Electricity) Grids for Smart Cities: Assessing Roles and Societal Impacts. Proceedings of the IEEE. 2018, 106(4): 613-625. doi: 10.1109/jproc.2018.2812212
9. Wei S, Qin S, Gao W, et al. The Research of Method of Malignant Load Identification Based on Wavelet Analysis. Physics Procedia. 2012, 25: 1070-1075. doi: 10.1016/j.phpro.2012.03.201
10. Bo L, Wenpeng L. Conceptual cloud solution architecture and application scenarios of power consumption data based on load disaggregation. Power System Technology. 2016, 40(3): 791-796.
11. Hart GW. Nonintrusive appliance load monitoring. Proceedings of the IEEE. 1992, 80(12): 1870-1891. doi: 10.1109/5.192069
12. Niu L, Jia H. Transient event detection algorithm for non-intrusive load monitoring. Automation of Electric Power Systems. 2011, 35(9): 30-35.
13. Wu X, Qi B, Han L, et al. Fast non-intrusive load identification algorithm for resident load based on template filtering. Automation of Electric Power Systems. 2017, 41(02): 135-141.
14. Wu H, Han X. Non-Intrusive Residential Load Decomposition Algorithm Solving Underdetermined Equations Based on Signal Sparsification. Power System Technology. 2017, 41(9): 3033-3040.
15. Zhou M, Song X, Tu J. Residential electricity consumption behavior analysis based on non-intrusive load monitoring. Power System Technology. 2018, 42(10): 3268-3274.
16. Ruyi L, Xiaohuan W, Meixuan H, et al. Application of RPROP neural network in nonintrusive load decomposition. Power System Protection and Control. 2016, 44(7): 55-61.
17. Batra N, Kelly J, Parson O, et al. NILMTK: an open source toolkit for non-intrusive load monitoring. In: Proceedings of the Fifth International Conference on Future Energy Systems; 11-13 June 2014; Cambridge, United Kingdom. pp. 265-276. doi: 10.1145/2602044.2602051
18. Figueiredo M, de Almeida A, Ribeiro B. Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems. Neurocomputing. 2012, 96: 66-73. doi: 10.1016/j.neucom.2011.10.037
19. Wu Z, Wang C, Peng W, et al. Non-intrusive load monitoring using factorial hidden markov model based on adaptive density peak clustering. Energy and Buildings. 2021, 244: 111025. doi: 10.1016/j.enbuild.2021.111025
20. Revuelta Herrero J, Lozano Murciego Á, López Barriuso A, et al. Non intrusive load monitoring (NILM): A state of the art. In: De la Prieta F, Vale Z, Antunes L, et al. (editors). Trends in Cyber-Physical Multi-Agent Systems. The PAAMS Collection - 15th International Conference, PAAMS 2017. Springer; 2017. Volume 619. pp. 125-138. doi: 10.1007/978-3-319-61578-3_12
21. Zoha A, Gluhak A, Nati M, et al. Low-power appliance monitoring using Factorial Hidden Markov Models. In: Proceedings of the 2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing; 2-5 April 2013; Melbourne, VIC, Australia. pp. 527-532. doi: 10.1109/issnip.2013.6529845
22. Wittmann FM, Lopez JC, Rider MJ. Nonintrusive Load Monitoring Algorithm Using Mixed-Integer Linear Programming. IEEE Transactions on Consumer Electronics. 2018, 64(2): 180-187. doi: 10.1109/tce.2018.2843292
23. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Available online: https://www.semanticscholar.org/paper/Attention-is-All-you-Need-Vaswani-Shazeer/204e3073870fae3d05bcbc2f6a8e263d9b72e776 (accessed on 28 February 2024).
24. Ridi A, Gisler C, Hennebert J. A survey on intrusive load monitoring for appliance recognition. In: Proceedings of the 2014 22nd International Conference on Pattern Recognition; 24-28 August 2014; Stockholm, Sweden. pp. 3702-3707. doi: 10.1109/ICPR.2014.636
25. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27-30 June 2016; Las Vegas, NV, USA. pp. 770-778.
26. Ba JL, Kiros JR, Hinton GE. Layer normalization. Available online: https://arxiv.org/abs/1607.06450 (accessed on 28 February 2024).
27. Kelly J, Knottenbelt W. 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; 4-5 November 2015; Seoul, South Korea. pp. 55-64. doi: 10.1145/2821650.2821672
28. Rafiq H, Zhang H, Li H, et al. Regularized LSTM based deep learning model: first step towards real-time non-intrusive load monitoring. In: Proceedings of the 2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE); 12-15 August 2018; Oshawa, ON, Canada. pp. 234-239. doi: 10.1109/SEGE.2018.8499519
29. Zhang C, Zhong M, Wang Z, et al. Sequence-to-point learning with neural networks for non-intrusive load monitoring. Proceedings of the AAAI Conference on Artificial Intelligence. 2018, 32(1). doi: 10.1609/aaai.v32i1.11873
30. Vinyals O, Bengio S, Kudlur M. Order matters: Sequence to sequence for sets. Available online: https://arxiv.org/abs/1511.06391 (accessed on 28 February 2024).