International Journal of Multiphysics
Volume 18, No. 3, 2024
ISSN: 1750-9548

Non-Invasive Load Decomposition Method Based on Multi-Scale TCN and Multi-Head Self-Attention Mechanism

Yan Zhang[1], Fei Li[1], Yang Xiao[1], Kai Li[2,*], Lei Xia[2], Huilei Tan[2]
[1]State Grid Beijing Changping Electric Power Supply Company, Beijing, China
[2]Wuxi Research institute of Applied Technologies, Tsinghua University, Wuxi, China
*Corresponding Author.

Abstract

As a branch of the development of smart grid, non-invasive load monitoring technology is important to promoting the information granularity of users' consumption behavior, improving the efficiency of power resource utilization, and promoting the sustainable development of smart power. For improve the efficiency of feature extraction and the accuracy of load decomposition, a non-invasive load decomposition method based on multi-scale TCN and multi-head self-attention mechanism is proposed in this paper. Firstly, expansive causal convolution of multi-scale TCN is used to expand the receptor field of convolutional kernel, and residual connection and batch normalization are added to heighten the quality and efficiency of extracting deep load features. Then, a multi-scale time information encoding and embedding method is proposed to enhance the model's capability to recognize the characteristics of electricity consumption behavior. Finally, the multi-head self-attention mechanism is used to extract important load features and historical key time point information, so as to capture the power series evolution pattern and complete load decomposition. In this paper, UKDALE and REDD residential power consumption data sets are used for training and testing. Results show that the model performs well, and the accuracy of part decomposition is enhanced compared with other existing methods.

Keywords: Non-invasive load decomposition, multi-scale TCN, time information embedding, self-attention mechanism

1. Introduction

In order to achieve the interactive quality between power grid and users and the informatization of power consumption management, intelligent distribution network puts forward new requirements for the granularity of power load monitoring [1-3]. For residential power consumption, electric-level fine-grained load operation information can be used to calculate demand-side response resources, guide residents to use electricity reasonably, and assist load modelling [4]. Non-intrusive Load Monitoring (NILM) is a key technique for acquiring intrusive load level operating information, a process that relies on a single-point measurement, characterized by power, voltage and current recorded by a total meter, to identify the operating state of an individual device and estimate its power consumption [5,6].

SNILM, based on traditional machine learning, consists basically of HMM [7-9], KNN [7], SVM [8], random forest [9] and other algorithms, such algorithms usually rely on relatively strict assumptions, and the description of load state change is not consistent with the actual operation characteristics of all electrical appliances. RNN[10,11] are a class architectures suitable for processing sequential data. For solve the long-term dependence problem of RNN, NILM problems are mostly based on LSTM networks [12,13] and GRU [14,15]. Convolutional Neural Networks (CNNS) also excel in sequence problems [16]. In 2017, Google proposed Transformer architecture [17,18], whose sequence modeling and information perception capabilities significantly surpass those of RNN and CNN, and the load decomposition effect has been effectively improved [19-21]. This kind of model can, to a certain extent, explore the relationship between a certain load event and the preceding and following events. However, since there is no global time label to describe long-term time sequence information, the scope of attention is limited to the internal sequence, and the capture of user behavior preferences is very limited, so the accuracy of model prediction is limited to some extent.

Based on the above analysis, a non-invasive load decomposition method based on multi-scale TCN and multi-head self-attention mechanism is proposed in this paper, mainly including the following contributions: (1) expansion causal convolution of multi-scale TCN is used to enlarge the receptor field of convolutional kernel. (2) a multi-scale time information encoding and embedding method is proposed to enhance the model's capability to recognize electricity consumption behavior features. (3) the multi-head self-attention is used to extract important load characteristics and historical key time point information, so as to capture the power series evolution pattern, complete the judgment of electrical switch and estimated power value, and realize load decomposition. In this paper, the Reference Energy Disaggregation Dataset (REDD) and UK Domestic Appliance-level Electricity (UKDALE), It is expected that the model generated by training performs well, and the accuracy of load decomposition can be enhanced compared with other existing methods.

2. A Non-Invasive Load Decomposition Model Based on Multi-Scale TCN and Self-Attention Mechanism

The network structure is shown in Figure 1. Using multi-scale TCN recognizes the partial feature information of different scales in the sequence, as well as considers the feature information of the sequence position, which makes the feature information more abundant. The multi-head self-attention mechanism is used to learn the context information of the power consumption sequence and take care the dependence between the distant data in the power consumption load data, for improve the precision of load decomposition. Finally, the feature vector is mapped to the electrical load sequence through the full connection layer.

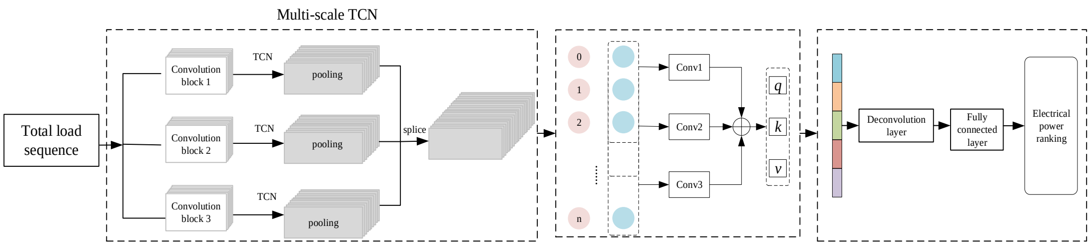

Figure 1 Model network structure diagram

2.1 Multi-scale TCN

The output of each layer in ordinary causal convolution is composed of the input of the corresponding neuron in the previous layer and the input of the previous position of the corresponding neuron. If the hierarchy of the network increases, the more layers there are between the input layer and the output layer, the more input information a neuron in the output layer corresponds to. Therefore, causal convolution requires longer filters or very deep network structures to achieve a sufficiently large receptive domain, but increasing network layers can also lead to problems such as gradient explosion. TCN is a variant of ordinary convolutional network, which combines extended convolutional network and causal convolutional network, and can be considered as extended causal convolution. Compared to ordinary convolutional networks, TCN improve the receptive field of convolutional kernels and also enhance the ability to process temporal data. TCN is a causal convolution that combines extended convolution, with a longer length and wider receptive field than ordinary causal convolution filters. The network structure of TCN can be seen in Figure 2. In this paper, extended convolution is introduced, which can handle the dependency of more complex classification data in the same case, and greatly improve the classification accuracy. The output of extended causal convolution is only related to the output information of the previous few moments and will not be affected by the information after the current moment. Due to the lengthening of the filter and the broadening of the receptive field of the convolutional kernel, the extended causal convolutional network can extract load characteristic information from time series data at a longer time. The receptive field of extended causal convolutional networks increases with the increase of network expansion rate, convolution kernel size, and network layers. Compared to regular convolutional networks, TCN have a larger receptive field and can extract more load features from sequence data with the same kernel size and network layers.

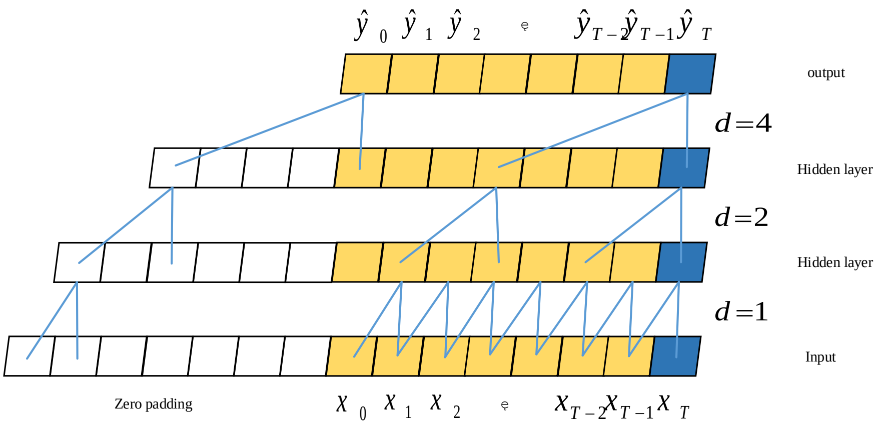

Figure 2 TCN network architecture diagram

The multi-scale TCN network structure is shown in Figure 3, which consists of 3 groups of parallel TCN modules. The TCN module performs feature learning and extraction on the input total power data, among which the 3 groups of parallel convolution kernels are (16, 1), (32, 1) and (64, 1) respectively. Finally, the features of the extracted power data from the 3 groups of TCN modules are spliced.

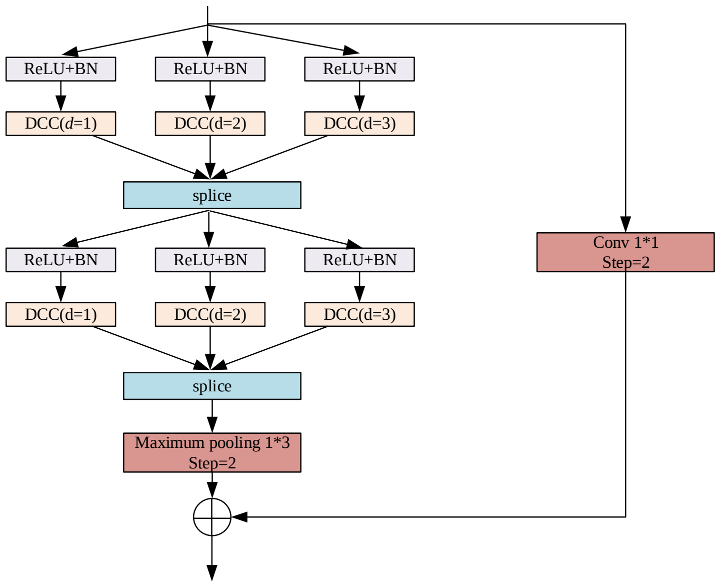

Figure 3 Multi scale TCN network architecture diagram

2.2 Temporal information embedding

The traditional load decomposition method only judges the state of a single electrical appliance by the aggregated power characteristics, and ignores the regularity of household users' electricity consumption to some extent. For example, dishwashers are usually used after lunch and dinner, microwave ovens and dishwashers are less likely to be used at noon on weekdays, and refrigerators are operated more frequently in hot seasons. The difference in appliance activity at different times of the day can be a strong indicator of appliance usage. However, due to the limited length of data fed into the model at one time, it is hard for the model to automatically capture the regular characteristics of a long time scale. Therefore, in this paper, the multi-scale time information is directly extracted and explicitly input into the model to enhance the sensitivity of the model to the law of electricity consumption. In this paper, the time information representation and embedding structure suitable for multi-scale TCN model are designed. The description method of time information can be seen in Figure 4. For ensure that the model accurately captures the time features most closely related to the household electricity consumption law and avoid the interference of redundant information, this paper extracts three key features for each time label: hours in 1 d, weeks in 1 week, and months in 1 a, and linearly codes these three features into values within the interval [-0.5, 0.5] respectively. For example, for the 2011-04-18 13:22:12 time label, the hours, weeks and months are 13, 1 and 4 respectively, so [0.065 2, -0.500 0, -0.3261] is given as a representation.

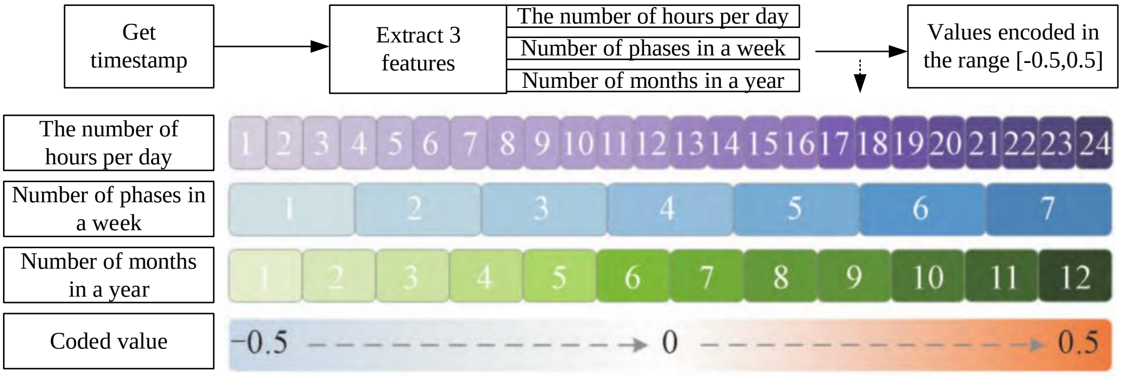

Figure 4 Schematic diagram of timestamp processing

2.3 Multi-head self-attention mechanism

The network structure of the Multi-head Self Attention mechanism can be seen in Figure 5. By stacking and scaling the attention structure of each layer in parallel, the model can effectively learn the signal features of different positions. Finally, the weighted summing result of the feature vectors of each layer of attention mechanism is taken as the network output. The output vector of the attention mechanism weight scoring function is:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

$$\begin{gathered} Hi = Attention(QW_Q^i, KW_K^i, VW_V^i) \\ MAS(Q,K,V) = concat(H_1, H_2, ..., H_h)W_0 \end{gathered} \tag{2}$$

where: K and V represent Key and Value pairs; Q is the target data; d_k is the dimension of Q, and the calculated value needs to satisfy Q=K=V. h is the number of attention heads and is the dimension of d_v V, the attention weight matrix. $W_0 \in R^{hdv \times \dim}, W_Q^i, W_K^i \in R^{\dim \times dk}, W_V^i \in R^{\dim \times dv}$.

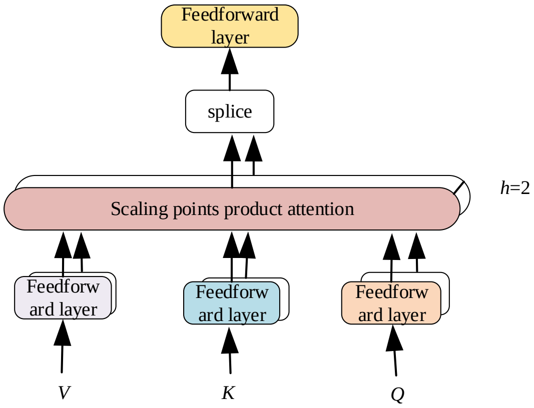

Figure 5 Multi-head selfattention mechanism network structure

2.4 Output Layer

The output layer part maps the power features extracted by the core decomposition layer to a single electrical power sequence of the same length as the total power sequence. The two fully connected layers convert the feature matrix into one-dimensional data of length L, and then adjust the output values according to the reasonable power range of each appliance to obtain the final predicted power.

3. Data Set and Data Processing

3.1 Data set selection

In this paper, REDD is the first and most widely used publicly available data set in the field of non-invasive part decomposition. It includes total electricity meters and load submeters of six U.S. homes, recorded for approximately 4 months. UKDALE is an open data set released by Imperial College London in 2014, and its data is very large, including the electrical level consumption data of five UK homes, of which the power collection time of House 1 is up to 3 years, and the other houses are several months.

3.1.1 Classification of training and test data

When the load decomposition of a house is carried out, the historical data of the electrical power consumption of the house is usually unknown, so the model is required to have a certain generalization ability for fresh samples, and the model trained with known house data can also decompose accurately on unknown houses.

In this paper, microwave oven, washing machine, dishwasher and refrigerator are selected as the research objects in REDD dataset, and microwave oven, washing machine, kettle and refrigerator are selected as the research objects in UKDALE dataset for cross-family train and testing. As shown in table 1, the training set and test set are divided.

Table 1 Selection of train and test set.

Data set	Training set housing number	Test set house number
REDD	2,3,4,5,6	1
UKDALE	1,3,4,5	2

3.1.2 Data set processing

Add status markers to the data set. According to the basic parameter setting of the appliance shown in table 2, if the power of the appliance is within the reasonable operating power range and the running time is greater than the minimum duration, the device is considered to be on, and the status value is marked as 1, otherwise it is marked as 0.

Table 2 Basic parameter settings of electric appliances.

Data set	Electric appliance	Appliance operation reasonable power range /W	Minimum duration of appliance operation /s	λ
REDD	Microwave oven	200~1800	12	5×10-2
	Washing machine	40~3500	1800	10-4
	Dish washer	50~1200	1800	5×10-4
	Kettle	500~3200	1800	5×10-4
	Frideg	50~400	60	10-4
UKDALE	Microwave oven	200~3 000	12	5×10-2
	Washing machine	300~2 500	1800	10-3
	Dish washer	2 000~3 100	12	10
	Kettle	500~3200	1800	5×10-4
	Frideg	50~300	60	10-4

Note: λ is the weight of the loss term

3.1.3 Evaluation criteria

The power data is standardized according to formula (3), where y is the original power data, y* is the normalized power data, μ is the sample mean and σ is the sample variance.

$$y* = \frac{y-\mu}{\sigma} \tag{3}$$

3.2 Example analysis

3.2.1 Experimental Environment

The environment of software platform is Windows10 operating system, Python3.7 (64-bit), TensorFlow-gpu2.5.0 and Keras2.5.0rc0 deep learning framework, and the integrated development environment is PyCharm Professional 2021.3. The experimental development environment is based on the open source NILMTK toolkit. The Pytorch deep learning framework was used in the Python programming platform to train and test the model.

3.2.2 Network parameters and loss function

The number, size and step size of convolutional kernel in input layer are 50, 5 and 1 respectively; the number of convolutional layers in TCN is 3, the number of filters is 50, the size of convolutional kernel is 5, and the expansion factors are set to 1, 2, and 4 respectively. The number of fully connected layer neurons was 128 and 1, respectively. In the training process, Batch_size was set to 128, the number of training iterations was 50, and the network parameter optimizer selected Adam. The loss function uses the MSE function:

$$\sigma_{MSE} = \frac{1}{T}\sum_{t=1}^{T}(\hat{x}_{it} - x_{it})^2 \tag{4}$$

Where: T is the number of sampling points of the power sequence; x dime is the decomposed power value of device i at time t; xit represents the actual of time t device i Power value.

3.2.3 Evaluation index

For measure the algorithm's ability to judge the load switching state and restore the load power consumption, it is necessary to select appropriate performance evaluation indexes. NILM task is not only a binary classification problem, but also a regression problem, so the performance of the algorithm can be evaluated by two kinds of indexes.

1) Classification performance evaluation index. The classification index in formula (5) can be used to evaluate the algorithm's capability to identify the electrical switch state, where NTP, NTN, NFP, NFN represents the number of sequence points in which the appliance is actually on and the decomposition result is also on, the appliance is actually on and the decomposition result is also on, the appliance is actually off and the decomposition result is on, the appliance is off and the decomposition result is also off, respectively. NP and NN represent the number of points where the appliance is actually turned on and off, respectively. Accuracy (A) is the most intuitive classification metric, but it is not suitable for unbalanced data sets and is not a good measure for situations where appliances are sparsely turned on. F score (F) takes into account the values of accuracy rate (P) and recall rate (R), so that β =1, and the F score at this time is the harmonic average of the two, called F1 score (F1).

$$
\begin{aligned}
A &= \frac{N_{TP} + N_{TN}}{N_P + N_N} \\
R &= \frac{N_{TP}}{N_{TP} + N_{FN}} \\
P &= \frac{N_{TP}}{N_{TP} + N_{FP}} \\
F &= (1+\beta^2) \times \frac{P \times R}{\beta^2 \times P + R} \\
F_1 &= 2 \times \frac{P \times R}{P + R}
\end{aligned}
\tag{5}
$$

2) Regression performance evaluation index. For evaluate the reduction effect of the model on the load power consumption, MAE and MRE, which are used in regression problems, are used to measure the decomposition performance of the algorithm. The values of MAE and MRE are expressed by EMA and EMR, respectively, and the formula:

$$
\begin{aligned}
E_{MA} &= \frac{1}{T} \sum_{t=1}^{T} \left| \hat{y}_t - y_t \right| \\
E_{MR} &= \frac{1}{T} \sum_{t=1}^{T} \frac{\hat{y}_t - y_t}{\max(\hat{y}_t, y_t)}
\end{aligned}
\tag{6}
$$

3.3 Experimental results

The proposed method was compared with 3 NILM algorithms, including Factorial Hidden Markov Model, FHMM) [8], bidirectional LSTM model [14], BERT4NILM model based on Transformer frame [19]. In this paper, two classification indexes of accuracy and F1 score are compared with two regression indexes of MAE and MRE. The performance indicators of the 7 decomposition methods on the REDD and UKDALE datasets are shown in Table 3 and Table 4.

In table 3 and table 4, it shown that on the two data sets, the five NILM methods based on deep learning models are overall better than the FHMM method. For most electrical appliances, the comprehensive performance of load decomposition is also superior.

Figure 6 shows the decomposition results of the first household electrical appliance data under different algorithm models. It shown that FHMM algorithm has poor decomposition performance for washing machines with complex running states. Although the part decomposition method proposed has some deviation in part of the running time, compared with other load decomposition methods, the model proposed can better decompose the power change of electrical appliances. In the washing machine such as the running state is more complicated equipment, the proposed model can better reflect the operation law of the equipment, for the microwave oven, refrigerator, dishwasher and other electrical appliances, the power change trend tracking effect is better, the decomposed power value is more accurate. It shown from Figure 6 that for electrical appliances with multiple running states similar to refrigerators, the model proposed has better decomposition effect than other models; In microwave ovens, kettles and dishwashers with few switching states and short running time, the decomposed power value can reach the actual peak value more, reflecting the actual operation of electrical equipment.

Table 3 Model performance on REDD dataset.

Item	Algorithm model	P ↑	F1 ↑	EMA ↓	EMR ↓
Washing machine	FHMM	0.903	0.173	38.47	0.091
	Bidirectional LSTM	0.993	0.124	35.63	0.020
	BERT4NILM	0.992	0.557	34.38	0.022
	Textual method	0.981	0.564	15.47	0.022
Microwave oven	FHMM	0.997	0.315	18.84	0.064
	Bidirectional LSTM	0.984	0.604	17.41	0.055
	BERT4NILM	0.989	0.476	17.58	0.057
	Textual method	0.990	0.559	16.53	0.056
Frideg	FHMM	0.793	0.518	31.65	0.894
	Bidirectional LSTM	0.573	0.174	43.74	0.956
	BERT4NILM	0.843	0.810	22.61	0.706
	Textual method	0.867	0.837	20.09	0.692
Kettle	FHMM	0.767	0.246	34.96	0.425
	Bidirectional LSTM	0.875	0.229	21.80	0.261
	BERT4NILM	0.943	0.503	11.47	0.197
	Textual method	0.966	0.664	7.04	0.179
Dish washer	FHMM	0.782	0.623	38.31	0.842
	Bidirectional LSTM	0.789	0.709	44.82	0.841
	BERT4NILM	0.770	0.638	46.91	0.849
	Textual method	0.873	0.793	29.99	0.807

Table 4 Model performance on UKDALE dataset.

Item	Algorithm model	P ↑	F1 ↑	EMA ↓	EMR ↓
Washing machine	FHMM	0.316	0.018	76.38	0.781
	Bidirectional LSTM	0.938	0.150	15.66	0.067
	BERT4NILM	0.995	0.325	6.98	0.040
	Textual method	0.996	0.299	3.16	0.011
Microwave oven	FHMM	0.970	0.029	7.83	0.016
	Bidirectional LSTM	0.995	0.060	6.55	0.014
	BERT4NILM	0.996	0.277	6.28	0.013
	Textual method	0.998	0.595	2.53	0.011
Frideg	FHMM	0.793	0.535	32.45	0.903
	Bidirectional LSTM	0.527	0.718	42.44	0.968
	BERT4NILM	0.855	0.885	23.71	0.739
	Textual method	0.897	0.903	20.89	0.643
Kettle	FHMM	0.989	0.418	23.98	0.009
	Bidirectional LSTM	0.994	0.531	21.26	0.007
	BERT4NILM	0.998	0.919	6.07	0.002
	Textual method	0.867	0.837	20.09	0.692
Dish washer	FHMM	0.767	0.246	34.96	0.425
	Bidirectional LSTM	0.875	0.229	21.80	0.261
	BERT4NILM	0.943	0.503	11.47	0.197
	Textual method	0.966	0.664	7.04	0.179

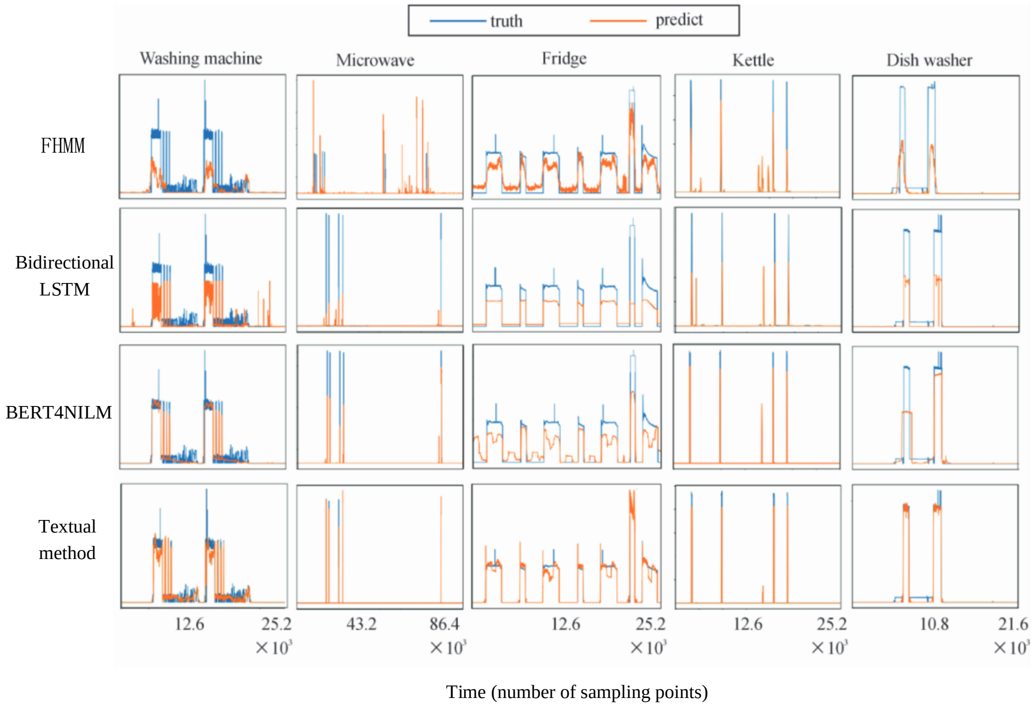

Figure 6 The results are decomposed by different algorithms in the UKDALE dataset

4. Conclusion

NILM technology is an important part of constructing a new power system. In this paper, firstly, expansive causal convolution of multi-scale TCN is used to expand the receptive field of the convolutional kernel, and residual connection and batch normalization are added to enhance the quality and efficiency of extracting deep load features. Then, a multi-scale time information encoding and embedding method is proposed to enhance the model's capability to recognize features of power consumption behavior. Finally, the multi-head self-attention mechanism is used to extract important load characteristics and historical key time point information, so as to capture the power series evolution pattern, complete the judgment of electrical switch and estimated power value, and realize load decomposition. In this paper, the REDD and UKDALE, Compared with other existing methods, the accuracy of part decomposition has been significantly improved. Moreover, the proposed algorithm has shown good results in cross-household tests, indicating that the algorithm has strong generalization ability in diverse residential electricity consumption scenarios. In the future, based on the existing research in this paper, it can be considered to further improve the model's ability to identify multi-mode devices with complex operating states, realize the monitoring of the running status of each electrical equipment of users, and provide detailed energy bills for the power grid company. At the same time, the application of NILM technology in detecting equipment faults and improving demand side response is also a direction worth studying in the future.

Acknowledgements

Funding Statement: This research was supported by the Science and technology projects of State Grid Corporation of China (Project Number: 520209230007).

References

[1] LIU W., HUANG Z., LI P., et al. Big data unified support platform for intelligent distribution grid System and framework. Transactions of China Electrotechnical Society, 2014, 29(S1): 486-491.

[2] YANG X., CAO Z. Q., ZHANG M. X., et al. Research on load decomposition of commercial buildings based on Attention-LSTM. Smart Power, 2020, 48(9):89-95.

[3] ZHAO J. H., DONG C. Y., WEN F. S., et al. Data science for energy systems: Theoretical techniques and prospects. Automation of Electric Power Systems, 2017, 41(4):1-11, 19.

[4] YAN X. F., DONG C. Y., WEN F. S., et al. Application of deep neural networks in non-invasive load decomposition. Automation of Electric Power Systems, 2019, 43(1):126-132, 167.

[5] YU D. W., LIU M., WANG Y. Q. Non-invasive household load decomposition based on GRNN and attention mechanism model. Smart Power, 2021, 49(3):74-79.

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

[7] KOLTER J. Z., JOHNSON M. J. REDD: a public data set for energy disaggregation research //Workshop on Data Mining Applications in Sustainability (SIGKDD). San Diego, USA, 2011, 25: 59-62.

[8] KIM H., MARWAH M., ARLITT M. F., et al. Unsupervised disaggregation of low frequency power measurements//Eleventh Siam International Conference on Data Mining. Mesa, USA, 2012:747-758.

[9] KOLTER J. Z., JAAKKOLA T. S. Approximate inference in additive factorial HMMs with application to energy disaggregation // International Conference on Artificial Intelligence and Statistics. La Palma, Spain, 2012:1472-1482.

[10] KELLY J., KNOTTENBELT W. Neural nilm: deep neural networks applied to energy disaggregation//2nd ACM international conference on embedded systems for energy-efficient built environments. Seoul, South Korea, 2015: 55-64.

[11] ELMAN J. L. Finding structure in time. Cognitive science, 1990, 14(2):179-211.

[12] GREFF K., SRIVASTAVA R. K., KOUTNÍK J., et al. LSTM: a search space odyssey. IEEE transactions on neural networks and learning systems, 2016, 28(10):2222-2232.

[13] XIA M., WANG K., SONG W., et al. Non-intrusive load disaggregation based on composite deep long short- term memory network. Expert Systems with Applications, 2020, 160(6):113669.

[14] RAFIQ H., ZHANG H. X., LI H. M., et al. Regularized LSTM based deep learning model: first step towards real- time non- intrusive load monitoring //2018 IEEE International Conference on Smart Energy Grid Engineering：Oshawa, Canada, 2018:234-239.

[15] KRYSTALAKOS O., NALMPANTIS C., VRAKAS D. Sliding window approach for online energy disaggregation using artificial neural networks//10th Hellenic Conference on Artificial Intelligence: Patras, Greece, 2018:1-6.

[16] EREN L., INCE T., KIRANYAZ S. A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. Journal of Signal Processing Systems, 2019, 91:179-189.

[17] ZHANG C. Y., ZHONG M. J., WANG Z. Z., et al. Sequence-topoint learning with neural networks for non- intrusive load monitoring //The 32nd AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018, 32(1):2604-2611.

[18] VASWANI A., SHAZEER N., PARMAR N., et al. Attention is all you need //The 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, USA, 2017:6000-6010.

[19] YUE Z., WITZIG C. R., JORDE D., et al. Bert4nilm: a bidirectional transformer model for non- intrusive load monitoring//5th International Workshop on Non-Intrusive Load Monitoring. [2020-11-18].

[20] SYKIOTIS S., KASELIMI M., DOULAMIS A., et al. Electricity: An efficient transformer for non-intrusive load monitoring. Sensors, 2022, 22(8):2926.

[21] YUE Z., ZENG H., KOU Z., et al. Efficient localness transformer for smart sensor-based energy disaggregation //2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS), Los Argeles, USA, 2022:141-148.