Improving Wireless Indoor Non-Intrusive Load Disaggregation Using Attention-based Deep Learning Networks

Qi Liu,[a,#] Jing Zhang,[a, #] Xiaodong Liu,[b] Yonghong Zhang[c]

[a]*School of Computer and Software; Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science & Technology, No. 219, Ningliu Road, Nanjing, 210044, China*

[b]*School of Computing, Edinburgh Napier University, 10 Colinton Road, Edinburgh, EH10 5DT, UK*

[c]*School of Automation, Nanjing University of Information Science & Technology, No. 219, Ningliu Road, Nanjing, 210044, China*

Elsevier use only: Received date here; revised date here; accepted date here

Abstract

One of the key functions of demand-side refinement management solutions is non-intrusive load monitoring (NILM), which has benefited from the growing interest in emerging technologies such as wireless communications and the Internet of Things. Currently, deep learning methods such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) are widely used for in-depth research on NILM. This paper investigates the role of attention mechanisms in the above two time-series deep learning models. Experiments show that the improved model is more than 10% more effective in indoor scenes, especially for typical household appliances such as refrigerators.

Keywords: CNN, LSTM, NILM, Attention, Load Management ;

[#] Both authors contributed equally to the work.

1. Introduction

Electricity supports the functioning of modern society, from food, clothing and housing to communication, entertainment and healthcare, electricity is an integral part of digital life [1]. The growing consumption of electrical energy has led to the development of the power industry, and the introduction of smart grids has led to an increase in the flexibility and stability of the power system. The development of power systems in the new era has gradually changed from scale expansion to efficiency enhancement, improving the sensing and control capabilities of the whole chain of power grid operation, focusing on model building and standard building of the underlying facilities, improving the quality of power grid data and strengthening data management have become urgent issues to be solved today [2].

Therefore, how to better access demand-side electricity data, further optimise the composition of power energy consumption, achieve digital transformation, improve energy utilisation, reduce carbon emissions and achieve sustainable development has become a major research direction today [3]. Advances in wireless indoor metering technology have opened up possibilities for power demand-side management in smart grid construction. Expanding from the simple metering functions of the past, smart meters now integrate wireless indoor monitoring functions such as remote meter reading and smart disconnection, providing the basis for further research into direct and indirect load management [4]. Indoor load management helps users to obtain real-time information on household energy consumption, thus prompting them to take the initiative to conserve energy, improve household energy utilization, reduce energy consumption and electricity load, and achieve significant economic and social benefits by reducing investment in new power plants and pollution of the atmosphere from primary energy sources [5]. It is estimated that more than 40% of the electricity consumed in the US is consumed in commercial buildings and residences, and load monitoring to help customers improve energy efficiency solutions can save more than 15% of energy consumption [6]. In addition, it can also improve the home life experience of home users and provide assistance for building intelligent integrated home services. Therefore, indoor load disaggregation has become a hot issue.

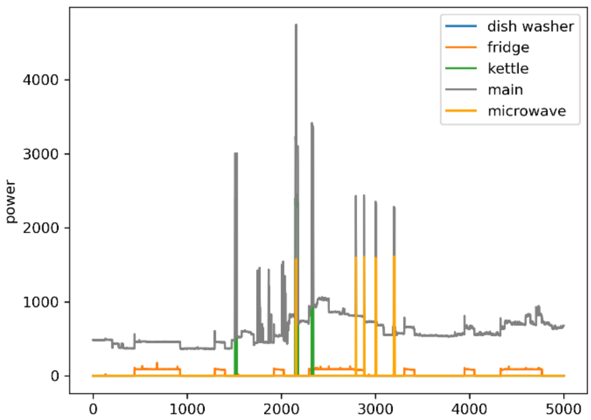


Fig. 1. Example of NILM.

Non-intrusive load monitoring aims to break down the total energy consumption data obtained from individual measuring devices to obtain the electricity consumption of each appliance, as shown in Fig. 1. It can effectively avoid the problems of high installation costs and inconvenient maintenance that make invasive monitoring difficult to promote, and provide a convenient monitoring solution for users [7]. Currently, clustering algorithms [8], hidden Markov models (HMM) [9], support vector machines (SVM) [10] and neural networks [11] have been tried to identify the energy consumption of different appliances. SVM are more accurate when there is a single load state, but they are computationally intensive and less accurate for continuously changing devices. Neural networks are relatively accurate, but they are susceptible to measurement noise and unknown devices. HMM solves the problem of discrete events and continuous value data, but it requires initialisation of the device state model with a priori knowledge, making it difficult to reason accurately.

With deep learning methods achieving better results in processing temporal data such as natural language, researchers have started to turn their attention to unsupervised deep learning algorithms [12]. Compared to supervised traditional statistical algorithms, unsupervised methods do not require manual extraction of domain knowledge; they treat all signals other than the target device as background noise and focus only on the features of the target signal [13]. It copes well with the situation of unknown devices that may exist in practical applications, but is less effective for multi-state device inference.

This paper investigates the impact of the Attention mechanism on load disaggregation, and adds the Attention module to common load disaggregation models to help the models focus effectively on key regions, from improving the disaggregation of multi-state devices. The work on NILM is briefly described in Section 2. Meanwhile, this paper adds the Atte mechanism to the bidirectional LSTM model and CNN model to effectively obtain different weights for the time series and focus on the target load active dimension, the details of which are shown in Section 3. The paper compares the LSTM and Seq2Point models, and the experimental results in Section IV show that the accuracy of the model's active state identification and energy consumption disaggregation are improved, especially for devices with complex states.

2. Related Work

Edge computing effectively mitigates latency issues, bringing a new leap forward in indoor wireless monitoring devices and the practical deployment of remote smart meter reading technology [14]. At the same time, cloud computing technology provides a safe and reliable solution for the collection and storage of massive amounts of power data [15]. With the widespread deployment of smart meters, publicly available datasets on home energy management provide the basis for research on indoor load management [16].

The traditional load disaggregation method sees it as a problem of solving for the power of individual devices using the total power obtained by superimposing the power of multiple devices A. It further decomposes different devices from the total energy consumption by fitting their power curves to model the different devices. Hidden Markov (HMM) has advantages in handling time-series data and is often used for modelling appliance power curves. Makonin et al. proposed a sparse HMM combined with an improved Viterbi algorithm to solve the NILM problem efficiently, but the method does not use time information and the disaggregation is poor. [17]. Kong et al. used the Hierarchical Hidden Markov Model (HHMM) framework to efficiently obtain temporal correlations using two Markov chains to achieve optimal identification of device switching state transitions, but the error in disaggregation accuracy is large and needs to be further optimized in combination with other models [18]. Bonfigli et al. combined active and reactive power characteristics based on the HMM model and validated it using the AMPds dataset, which improved performance by more than 10% compared to traditional methods, but the model effect began to decline as the number of devices increased [19]. The HMM model effectively solves the problem of discrete-event and continuous-valued data, but the model is difficult to infer accurately and is susceptible to noise.

Due to the emphasis on research in the area of load disaggregation by government departments, the data content contained in the power dataset is becoming more and more refined, which facilitates the application of supervised deep learning methods. Fabrizio et al. used the current from the BLUED dataset as an input quantity and performed simultaneous event detection and classification through three layers of convolution to improve computational efficiency, and achieved an accuracy rate of over 85% [20]. Pascal et al. used contextual information to extract device features forward for effective classification, and the results showed that the improved algorithm improved for four different types of devices, especially continuous devices [21]. The analysis was conducted [22]. Although deep learning methods have achieved better results in load disaggregation, there is still room for improvement.

In recent years, the attention mechanism's use of different weight assignments makes the model focus on key features and has achieved better results in computer vision and natural language processing, so researchers have also started to apply it to other fields with hierarchical representation learning [23]. Ji et al. combined the long and short-term memory model with the attention mechanism to effectively transfer visual attention to achieve image caption annotation [24]. Zhao et al. applied the power of hierarchical attention mechanism to the field of emotion recognition and effectively identified depression text problems through speech information under semi-supervised learning methods [25].

3. An Attention-based Time-series Deep Learning Model

In the traditional approach, the total power can be regarded as the accumulation of the energy consumption of several devices. The load disaggregation problem can then be considered as a reverse process, i.e. the process of solving the power sequence of individual devices for a known sum, which can be expressed as the following equation:

$$E(x) = \sum E_i(x) + e(x) \quad (1)$$

where $E(x)$ indicates the total power obtained from the measurement, $E_i(x)$ indicates the power of the individual appliance and $e(x)$ indicates the measurement noise value.

However, it is different in the deep learning approach, where the model treats values other than the target power sequence as noise and the whole computation process can be seen as a denoising process. Therefore, the influence of other devices can be ignored and only the features of the target signal need to be focused on. The equation can be expressed as follow:

$$E(x) = P_i(x) + e(x) \quad (1)$$

where E(x) denotes the total power obtained from the measurement, $P_i(x)$ denotes the power of an individual device and $e(x)$ denotes the measurement noise value.

3.1. The attention mechanism

For time-series data modelling, CNNs and LSTMs have been used, but conventional methods can be neglected in the process of acquiring the main features of the device. The attention mechanism, which mimics biological observation behaviour, can align internal experience with external senses to observe key regions more finely, with less complexity and parameters, and increase the weight of key regions through the probability of attention, thus allowing the model to achieve a better training effect [26]. The unit structure is shown in Fig. 2.

The attention mechanism establishes a mapping of the learning target q to the dictionary k by means of a

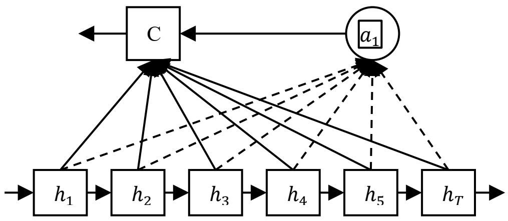


Fig. 2. The attention mechanism unit

weight matrix. The input is first subjected to a transpose, then softmaxed for each dimension to obtain the attention weights for each dimension, and finally merged into the attention weights for the individual features, which are given by the following equation:

$$e(q,k) = q^T k \quad (2)$$

$$\alpha_t = softmax\big(e(q,k)\big) = \frac{\exp\left(e(q,k_i)\right)}{\sum \exp\left(e(q,k_i)\right)} \quad (3)$$

$$Attention = (q,k,v) = \sum a_t v \quad (4)$$

The attention mechanism used in this paper is shown in Figure 3, where the model utilises a Permute layer to swap the data in dimensions 2 and 3. The weights of the features in each TIME_STEP are then calculated. Finally, the weighted output is multiplied with the input to obtain the weighted sequence.

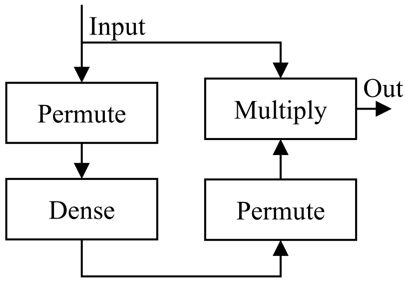


Fig. 3. The attention mechanism unit.

3.2. The attention-based LSTM network

Recurrent neural networks are chosen as the main architecture in this paper because they have been shown to perform well in time series. Whereas NILM is the network input is a period of power consuming time series data and has achieved better results on LSTM models.

LSTM models are special RNN models that use cells with memory functions to selectively control the input of information [27]. The attention-based bidirectional recurrent LSTM network adds the ATT module to the original model architecture, the detailed structure of which is shown in Fig. 3(a). the LSTM neurons ensure that the network input parameters can be preserved for a long time and have good performance in long-term learning [28]. The output of a two-way LSTM prediction is determined by a number of inputs in front of it and a number of inputs behind it, and its prediction will be more accurate compared to the LSTM model. However, due to its relative computational complexity, high computational cost during training and large disaggregation time cost, it needs further improvement in practical applications.

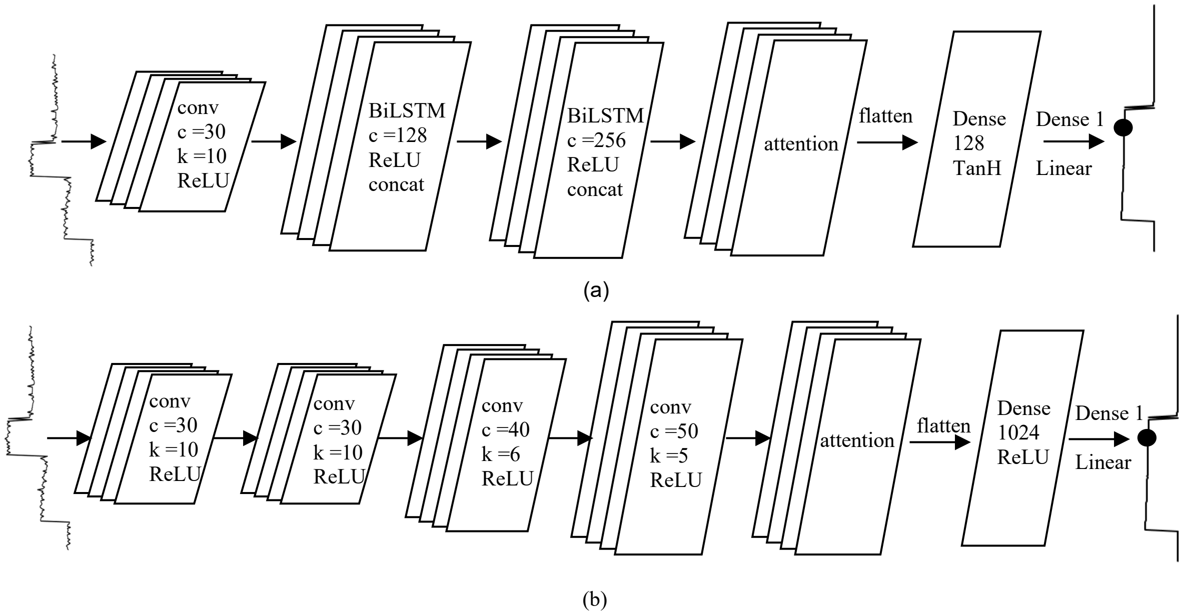


Fig. 4. (a) Architecture of attention-based LSTM network; (b) Attention-based Seq2P network architecture.

3.3. The attention-based Seq2P network

CNNs are selected as the main network structure in this paper because of their simple method structure, less training parameters and computation time than recurrent neural networks, and better training results. Also, in order to accommodate the time series data as input, the CNN network utilises a sliding window of length 100 as input for one-dimensional convolution.

Convolutional neural networks can perceive local information to obtain further global information, effectively reducing the number of parameters [29]. The Attention-based seq2point network made use of four CNNs for feature extraction, while the dropout layers introduced after each layer hides different neurons randomly in each cycle, thus avoiding the phenomenon of overfitting and enhancing the generalisation ability of the model. The additional attention layer is responsible for calculating the weight parameter matrix, and finally the final output is obtained by fully connected calculation. The specific structure of the model is shown in Fig. 3(b).

Before training and prediction, we normalised the data to speed up the network training. Compared to LSTM networks, convolutional neural networks are faster, taking only about 60us per sample compared to about 180us for LSTM networks. This is due to the fact that CNNs do not have the problem of back-and-forth dependency and can operate efficiently in parallel. In practical deployments, efficient and lightweight disaggregation algorithms are beneficial for dissemination and application, and therefore the computation speed needs to be optimised while improving the accuracy.

4. Experiments

4.1. The dataset

This paper uses the REDD[30] dataset from the USA and the UKDALE[31] dataset for training, REDD contains data from 6 households, from which 4 devices were selected for evaluation, and UKDALE contains data from 5 households, from which 5 devices were selected for evaluation. The sampling frequency of REDD was 1s and that of UKDALE was 6s, and the specific devices are shown in Table 1.

Table 1

Devices selected in REDD and UKDALE

Device	REDD	UKDALE
Kettle	-	√
Fridge	√	√
Microwave	√	√
Dish washer	√	√
Washing machine	√	√

Prior to training, the data were first pre-processed by down sampling the data to 6S and filling the sequence with 0's with values of NAN, followed by normalisation.

4.2. The metrics

This paper evaluates the model performance in terms of both the identification of device switching states and the accuracy of the disaggregation values. The following indicators were used as model assessments.

$$Rec = \frac{TP}{TP+FN} \tag{3}$$

$$Pre = \frac{TP}{TP+FP} \tag{4}$$

$$F1 = 2 * \frac{Pre*Rec}{Pre+Rec} \tag{5}$$

$$Acc = 2 * \frac{Pre*Rec}{Pre+Rec} \tag{6}$$

$$MAE = \frac{1}{T}\sum_{T=1}^{T} |x_t' - x_t| \tag{7}$$

$$RETE = \frac{|P' - P|}{\max(P', P)} \tag{8}$$

In this paper, the device on/off state is regarded as a binary classification problem, and the predicted and true values greater than the turn-on threshold are marked as 1, and those less than the turn-on threshold are marked as 0. The turn-on threshold is obtained by ground.on_power_threshold(). TP is the quantity where both are 1 and TN is the quantity where both are 0. When the true value is marked as 0 and the predicted value is marked as 1, it is counted as FP and the rest as FN. Where Rec represents the ratio of correctly predicted samples and Pre represents the accuracy of the device recognition. In addition, $F1$ is used to consider the performance of Rec and Pre together.

Considering the problem of the accuracy of the predicted values that can be returned in practical applications, Mean Absolute Error (MAE) and relative error in total energy ($RETE$) are introduced in this paper to assess the error between the predicted and true values. These two indicators are used to assess the accuracy of the model disaggregation values; the smaller the value, the smaller the error.

4.3. Results

The attention-based LSTM network has F set to 16, K set to 4, and liner as the activation function, and the two LSTM layers are set differently at 128 and 256. the attention-based Seq2P network has a window length of 100 and the convolutional layer activation function is ReLU. The results of the experiment are shown in the table below.

Table 2

Disaggregation performance for fridge

Methods	Rec↑	Pre↑	Accuracy↑	F1↑	RETE↓	MAE↓
RNN	**0.9632**	0.3678	0.4969	0.5324	0.1234	38.96
S2P	0.8695	0.7729	0.8855	0.8184	0.1058	18.36
Att-LSTM (this paper)	0.9374	0.4304	0.6127	0.59	0.1989	36.37
Att-S2P (this paper)	0.8623	**0.86**	**0.9175**	**0.8612**	**0.0543**	**14.66**

Table 3

Disaggregation performance for microwave

Methods	Rec↑	Pre↑	Accuracy↑	F1↑	RETE↓	MAE↓
RNN	0.2087	0.2192	0.8985	0.2138	0.4694	22.81
S2P	0.2071	0.0061	0.9007	0.2162	0.0082	16.91
Att-LSTM (this paper)	**0.2132**	0.2103	0.895	0.2117	0.2615	25.36
Att-S2P (this paper)	0.2016	**0.5374**	**0.9345**	**0.2923**	**0.0042**	**12.46**

Table 4

Disaggregation performance for dish wash

Methods	Rec↑	Pre↑	Accuracy↑	F1↑	RETE↓	MAE↓
RNN	0.9356	0.0982	0.1973	0.1778	**0.1130**	43.03
S2P	**0.99**	0.0921	0.0921	0.1687	0.5052	66.79
Att-LSTM (this paper)	0.9701	**0.0943**	0.1327	0.1718	0.4144	**36.6**
Att-S2P (this paper)	0.9478	0.0946	**0.1596**	**0.1720**	0.1492	48.41

Table 5

Disaggregation performance for washing machine

Methods	Rec↑	Pre↑	Accuracy↑	F1↑	RETE↓	MAE↓
RNN	0.6160	0.7527	0.9837	0.6775	0.2853	21.89
S2P	0.6824	0.6908	0.9820	0.6766	0.8778	47.11
Att-LSTM (this paper)	0.6096	**0.7741**	**0.9843**	0.6820	**0.0969**	28.87
Att-S2P (this paper)	**0.7277**	0.6442	0.9814	**0.6834**	0.1506	**19.46**

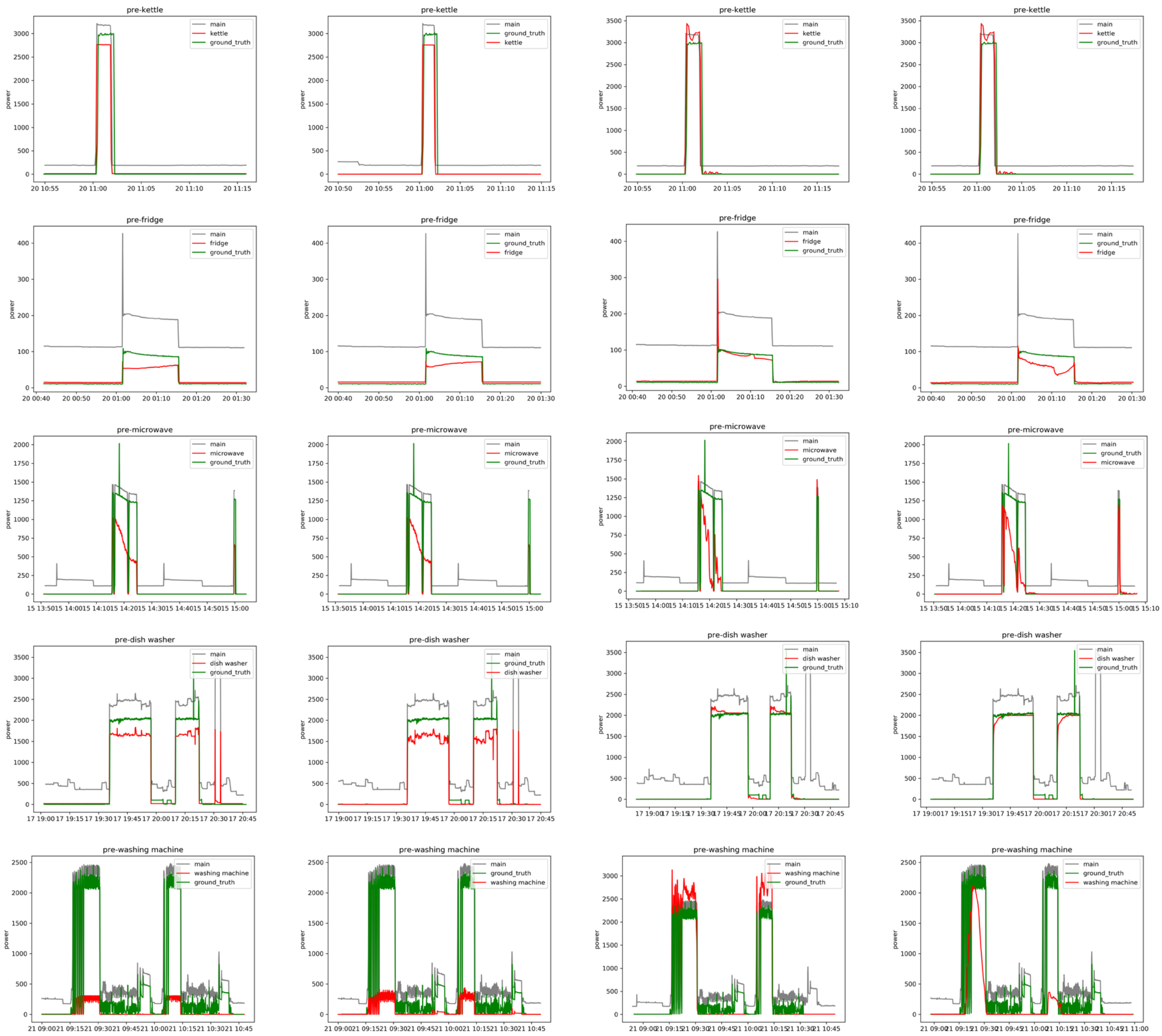
Table 6

Disaggregation performance for kettle

Methods	Rec↑	Pre↑	Accuracy↑	F1↑	RETE↓	MAE↓
RNN	0.1456	0.1936	0.7497	0.1662	0.5569	25.16
S2P	0.1454	**0.6600**	0.81	0.2382	**0.0617**	24.94
Att-LSTM (this paper)	0.0482	0.3298	0.7972	**0.3917**	0.1549	**17.19**
Att-S2P (this paper)	**0.1981**	0.3525	**0.9633**	0.2536	0.3512	19.86

As we can see from the results of the experiment, the best results were boosted by the fridge and kettle. The results for washing machine, dishwasher and microwave oven are not so obvious. It can be seen that the attention mechanism can effectively improve the training effect of the model, while the reduction of MAE and RETE values can show that the accuracy of energy consumption prediction has been improved.

In the prediction of refrigerators, the improved model improved by more than 10%, but the improved Att-LSTM still could not catch up with the S2P. There was a significant improvement in the prediction of washing machines, which may be due to the high number of feature channels of washing machine equipment, of which the attention module effectively focuses on the key areas. In addition, the improved Att-S2P outperformed the best GRU in terms of F1



Att-LSTM	RNN	ATT-S2P	S2P

Fig.5. Some example disaggregation result on UK-DALE.

and MAE, and in the case of the kettle, the model improved by 20%, which means that the attention module can effectively extract the hidden features of the more obvious devices. The Att-S2P effect was much improved on the prediction of microwave ovens, probably due to the high number of feature channels of microwave oven devices, of which the attention module effectively focused on the key areas. In addition, in the other two equipment energy consumption predictions, the model did not improve as well, probably because the characteristic pattern of equipment with complex operating states is not obvious, and the simple attention module is not well suited to the identification of multi-state equipment.

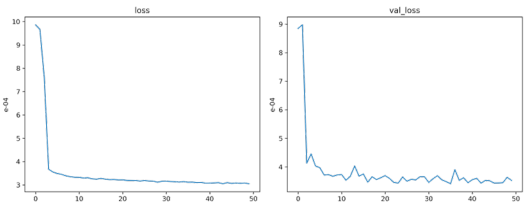


Fig.6. Outputs of loss and val_loss of 50 epoches.

Fig. 5 shows some example disaggregation results on UK-DALE. From the figure we can see that the attention mechanism improves better than the ATT-LSTM on the S2p model, this is due to the fact that the LSTM model itself can acquire temporal features very well, so the improvement is not so obvious. At the same time, it can be seen that the ATT-S2P model is more effective for devices with fixed patterns such as washing machines, dishwashers and refrigerators. The changes of loss and val_loss of 50 epoches are shown in Fig. 6.

5. Conclusion

This paper studies the role of attention mechanism in the field of load disaggregation. The experiment shows that the attention module helps to improve the model effect, and the model effect is improved by more than 10%. At the same time, for S2p model, ATT mechanism reduces the disaggregation error of equipment with fixed operation cycle. Although att-seq2p model has many layers, its calculation is relatively simple and its training speed is faster.

Considering the characteristics of different electrical appliances and users can improve the accuracy of load disaggregation algorithm. Next, the author will further improve the attention based deep learning networks proposed in this paper, so that it can more accurately reflect the characteristics of electrical appliances. Because the performance of the improved S2p model is better than the improved LSTM model, the author's next work will be carried out in the S2P model.

Acknowledgment

This work has received funding from the Key Laboratory Foundation of National Defence Technology under Grant 61424010208, National Natural Science Foundation of China (No. 41911530242 and 41975142), 5150 Spring Specialists (05492018012 and 05762018039), Major Program of the National Social Science Fund of China (Grant No. 17ZDA092), 333 High-Level Talent Cultivation Project of Jiangsu Province (BRA2018332), Royal Society of Edinburgh, UK and China Natural Science Foundation Council (RSE Reference: 62967_Liu_2018_2) under their Joint International Projects funding scheme and basic Research Programs (Natural Science Foundation) of Jiangsu Province (BK20191398 and BK20180794).

References

[1] O. A. Páramo, J. C. Rivera and G. A. López, "Effects on Electricity Markets of a Demand Response Model Based on Day Ahead Real Time Prices: Application to the Colombian Case," in IEEE Latin America Transactions, vol. 16, no. 5, pp. 1416-1423, 2018, https://doi.org/10.1109/TLA.2018.8408436

[2] X. He, Q. Ai, R. C. Qiu, W. Huang, L. Piao and H. Liu, "A Big Data Architecture Design for Smart Grids Based on Random Matrix Theory," in IEEE Transactions on Smart Grid, vol. 8, no. 2, pp. 674-686, 2017, https://doi.org/10.1109/TSG.2015.2445828.

[3] H. Yang, J. Zhang, J. Qiu, S. Zhang, M. Lai and Z. Y. Dong, "A Practical Pricing Approach to Smart Grid Demand Response Based on Load Classification," in IEEE Transactions on Smart Grid, vol. 9, no. 1, pp. 179-190, 2018, https://doi.org/10.1109/TSG.2016.2547883.

[4] M. U. Saleem, M. R. Usman and M. Shakir, "Design, Implementation, and Deployment of an IoT Based Smart

Energy Management System," IEEE Access, vol. 9, pp. 59649-59664, 2021, https://doi.org/10.1109/ACCESS.2021.3070960.

[5] S. Moon and J. Lee, "Multi-Residential Demand Response Scheduling With Multi-Class Appliances in Smart Grid," in IEEE Transactions on Smart Grid, vol. 9, no. 4, pp. 2518-2528, 2018, https://doi.org/10.1109/TSG.2016.2614546.

[6] C. Nalmpantis and D. Vrakas, "Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparation," Artificial Intelligence Review, vol. 2019, no. 52, pp. 217-243, 2019, https://doi.org/10.1007/s10462-018-9613-7.

[7] F. M. Wittmann, J. C. López and M. J. Rider, "Nonintrusive Load Monitoring Algorithm Using Mixed-Integer Linear Programming," in IEEE Transactions on Consumer Electronics, vol. 64, no. 2, pp. 180-187, 2018, https://doi.org/10.1109/TCE.2018.2843292.

[8] H. Liu, Q. Zou and Z. Zhang, "Energy Disaggregation of Appliances Consumptions Using HAM Approach," in IEEE Access, vol. 7, no. 1 pp. 185977-185990, 2019, https://doi.org/10.1109/ACCESS.2019.2960465.

[9] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo and Y. Xu, "Improving Nonintrusive Load Monitoring Efficiency via a Hybrid Programing Method," in IEEE Transactions on Industrial Informatics, vol. 12, no. 6, pp. 2148-2157, 2016, https://doi.org/10.1109/TII.2016.2590359.

[10] K. T. Chui, M. D. Lytras, A. Visvizi, "Energy Sustainability in Smart Cities: Artificial Intelligence, Smart Monitoring, and Optimization of Energy Consumption," Energies, vol. 11, no. 11, pp. 2869-2888, 2018, https://doi.org/10.3390/en11112869

[11] D. Yang, X. Gao, L. Kong, Y. Pang and B. Zhou, "An Event-Driven Convolutional Neural Architecture for Non-Intrusive Load Monitoring of Residential Appliance," in IEEE Transactions on Consumer Electronics, vol. 66, no. 2, pp. 173-182, 2020, https://doi.org/10.1109/TCE.2020.2977964.

[12] L. K. Şenel, İ. Utlu, V. Yücesoy, A. Koç and T. Çukur, "Semantic Structure and Interpretability of Word Embeddings," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 10, pp. 1769-1779, 2018, https://doi.org/10.1109/TASLP.2018.2837384.

[13] A. U. Rehman, T. T. Lie, B. Vallès and S. R. Tito, "Event-Detection Algorithms for Low Sampling Nonintrusive Load Monitoring Systems Based on Low Complexity Statistical Features," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 3, pp. 751-759, 2020, https://doi.org/10.1109/TIM.2019.2904351.

[14] D. Gupta, S. Rani, S. H. Ahmed, S. Verma, M. F. Ijaz, and J. Shafi, "Edge Caching Based on Collaborative Filtering for Heterogeneous ICN-IoT Applications," Sensors, vol. 21, no. 16, pp. 5491, 2021. https://doi.org/10.3390/s21165491.

[15] A. W. Khan et al., "Analyzing and Evaluating Critical Challenges and Practices for Software Vendor Organizations to Secure Big Data on Cloud Computing: An AHP-Based Systematic Approach," in IEEE Access, vol. 9, pp. 107309-107332, 2021, https://doi.org/10.1109/ACCESS.2021.3100287.

[16] M. Sun, F. M. Nakoty, Q. Liu, X. Liu, Y. Yang and T. Shen, "Non-Intrusive Load Monitoring System Framework and Load Disaggregation Algorithms: A Survey," 2019 International Conference on Advanced Mechatronic Systems (ICAMechS), 2019, pp. 284-288, https://doi.org/10.1109/ICAMechS. 2019. 8861646.

[17] S. Makonin, F. Popowich, I. V. Bajić, B. Gill and L. Bartram, "Exploiting HMM Sparsity to Perform Online Real-Time Nonintrusive Load Monitoring," in IEEE Transactions on Smart Grid, vol. 7, no. 6, pp. 2575-2585, 2016, https://doi.org/10.1109/TSG.2015.2494592.

[18] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. H. Zhao and F. J. Luo, "A Hierarchical Hidden Markov Model Framework for Home Appliance Modeling," in IEEE Transactions on Smart Grid, vol. 9, no. 4, pp. 3079-3090, 2018, https://doi.org/10.1109/TSG.2016.2626389.

[19] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, F. Piazza, "Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models, " APPLIED ENERGY -BARKING THEN OXFORD, vol. 208, no. 15, pp. 1590-1607, 2017, https://doi.org/10.1016/j.apenergy.2017.08.203.

[20] F. Ciancetta, G. Bucci, E. Fiorucci, S. Mari and A. Fioravanti, "A New Convolutional Neural Network-Based System for NILM Applications," in IEEE Transactions on Instrumentation and Measurement, vol. 70, no.1, pp. 1-12, 2021, https://doi.org/10.1109/TIM.2020.3035193.

[21] P. Schirmer, I. Mpora, A. Sheikhakbari, "Robust energy disaggregation using appliance-specific temporal contextual information," Journal on Advances in Signal Processing, , vol. 2020, no. 1, 2020, https://doi.org/10.1186/s13634-020-0664-y.

[22] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis and A. Doulamis, "Context Aware Energy Disaggregation Using Adaptive Bidirectional LSTM Models," in IEEE Transactions on Smart Grid, vol. 11, no. 4, pp. 3054-3067, 2020, https://doi.org/10.1109/TSG.2020.2974347.

[23] W. Du, Y. Wang and Y. Qiao, "Recurrent Spatial-Temporal Attention Network for Action Recognition in Videos," in IEEE Transactions on Image Processing, vol. 27, no. 3, pp. 1347-1360, 2018, https://doi.org/10.1109/TIP.2017.2778563.

[24] J. Ji, C. Xu, X. Zhang, B. Wang and X. Song, "Spatio-Temporal Memory Attention for Image Captioning," in IEEE Transactions on Image Processing, vol. 29, no.1 , pp. 7615-7628, 2020, https://doi.org/10.1109/TIP.2020.3004729.

[25] Z. Zhao, Z. Bao, Z. Zhang, J. Deng, N. Cummins, H. Wang et al., "Automatic Assessment of Depression From Speech via a Hierarchical Attention Transfer Network and Attention Autoencoders," in IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 2, pp. 423-434, 2020, https://doi.org/10.1109/JSTSP.2019.2955012.

[26] Y. Zhu, C. Zhao, H. Guo, J. Wang, X. Zhao and H. Lu, "Attention CoupleNet: Fully Convolutional Attention Coupling Network for Object Detection," in IEEE Transactions on Image Processing, vol. 28, no. 1, pp. 113-126, 2019, https://doi.org/10.1109/TIP.2018.2865280.

[27] B. Brattoli, U. Büchler, A. Wahl, M. E. Schwab and B. Ommer, "LSTM Self-Supervision for Detailed Behavior Analysis," 2017 IEEE Conference on Computer Vision and Pattern

Recognition (CVPR), 2017, pp. 3747-3756, https://doi.org/10.1109/CVPR.2017.399.

[28] T. Ergen and S. S. Kozat, "Online Training of LSTM Networks in Distributed Systems for Variable Length Data Sequences," in IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 10, pp. 5159-5165, 2018, https://doi.org/10.1109/TNNLS.2017.2770179.

[29] J. Cao, Y. Pang and X. Li, "Learning Multilayer Channel Features for Pedestrian Detection," in IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3210-3220, 2017, https://doi.org/10.1109/TIP.2017.2694224.

[30] J. Z. Kolter and M. J. Johnson, “REDD: A public data set for energy disaggregation research,” in Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA, vol. 25, pp. 59–62, 2011, http://redd.csail.mit.edu.

[31] Kelly, J., Knottenbelt, W.: UK-DALE: a dataset recording UK domestic appliance-level electricity demand and whole-house demand. arXiv, April 2014, https://doi.org/10.1038/sdata.2015.7.