Attention-guided Temporal Convolutional Network for Non-intrusive Load Monitoring

1st Huamin Ren
School of Economics, Innovation and Technology
Kristiania University College
Oslo, Norway
huamin.ren@kristiania.no

2nd Xiaomeng Su
Dept. of Computer Science
Norwegian University of Science and Technology
Trondheim, Norway
xiaomeng.su@ntnu.no

3rd Robert Jenssen
Dept. of Physics and Technology
UiT The Arctic University of Norway
Tromsø, Norway
robert.jenssen@uit.no

4th Jingyue Li
Dept. of Computer Science
Norwegian University of Science and Technology
Trondheim, Norway
jingyue.li@ntnu.no

5th Stian Normann Anfinsen
NORCE Norwegian Research Centre
UiT The Arctic University of Norway
Tromsø, Norway
stian.normann.anfinsen@uit.no

***Abstract*—With the prevalence of smart meter infrastructure, data analysis on consumer side becomes more and more important in smart grid systems. One of the fundamental tasks is to disaggregate users' total consumption into appliance-wise values. It has been well noted that encoding of temporal dependency is a key issue for successful modelling of the relations between the total consumption and its decomposed consumption on an appliance historically, and therefore has been implemented in many state-of-the-art models. However, how to encode the varied long-term and short-term dependency coming from different appliances is yet an open and under-addressed question. In this paper, we propose an Attention-guided Temporal Convolutional Network (ATCN), which generates different temporal residual blocks and provides an attention mechanism to indicate the importance of those blocks with respect to the appliance. Ultimately, we aim to address these two questions: i) How to employ both long-term and short-term temporal dependency to better disaggregate future loads while maintaining an affordable memory cost? ii) How to employ attention during the training of an appliance to obtain a better representation of the consumption pattern? We have demonstrated the effectiveness of our approach through comprehensive experiments and show that our proposed ATCN model achieves state-of-the-art performance, particularly on multi-status appliances that are normally hard to cope with regarding disaggregation accuracy and generalization capability.**

***Index Terms*—energy disaggregation, non-intrusive load monitoring, deep learning, temporal convolutional network, attention model**

I. Introduction

Non-Intrusive Load Monitoring (NILM), also referred to as energy disaggregation, aims to disaggregate the power consumption of a customer as a whole into detailed appliance-level consumption [1]. It has become one of the key tools to make effective use of the emerging smart meter infrastructure for the benefit of energy customers and producers, with great potential in applications such as energy awareness, energy conservation, and identification of controllable loads [2].

NILM has been framed historically both as classification and regression problems. Our paper is treating NILM as a regression problem, i.e, to estimate the consumption of individual appliances from the mains signal. In order to capture all distinct consumption patterns from all types of appliances, NILM algorithms tend to adopt a training dataset with a long time span (as long as memory permits) and attempt to learn temporal dependencies for each appliance. The trend is that recent work tends to utilize a range of deep neural network architectures, such as encoder-decoder networks, long short-term memory (LSTM) networks, bi-directional, sequence-to-sequence, and sequence-to-point [3] [4] [5] based prediction algorithms and their variants, including the very recent Bitcn-NILM algorithm [6], which combines sequence-to-point with bidirectional dilated convolution network. The key challenges of the prediction strategy are these: if the time window is too small, essential dependencies cannot be learned, e.g. if an appliance has a cyclic consumption pattern and the time window does not cover a full period. However, if it is too large, the efficiency of the scheme can significantly degrade, since loading long historical data burdens the memory requirement. Additionally, it also requires a much longer prediction time, and therefore cannot meet the needs of real applications.

Remark that different appliances exhibit vastly different temporal dependencies. As an example, we present in Figure 1 the daily power consumption of a household from the REDD dataset (shown in black color) along with its corresponding appliance-level consumption. Two important findings can be seen: 1) both local neighbors and far-away neighbors in disjoint time windows can together help with the disaggregation of appliance-level consumption. 2) for some appliances, local neighbors are most important for prediction of future consumption (e.g., fridge in Fig.1); for others, far-away points play a more important role (e.g., microwave and washer dryer in Fig.1). In other words, the relevant dependency ranges are specific to each appliance and should be adapted accordingly.

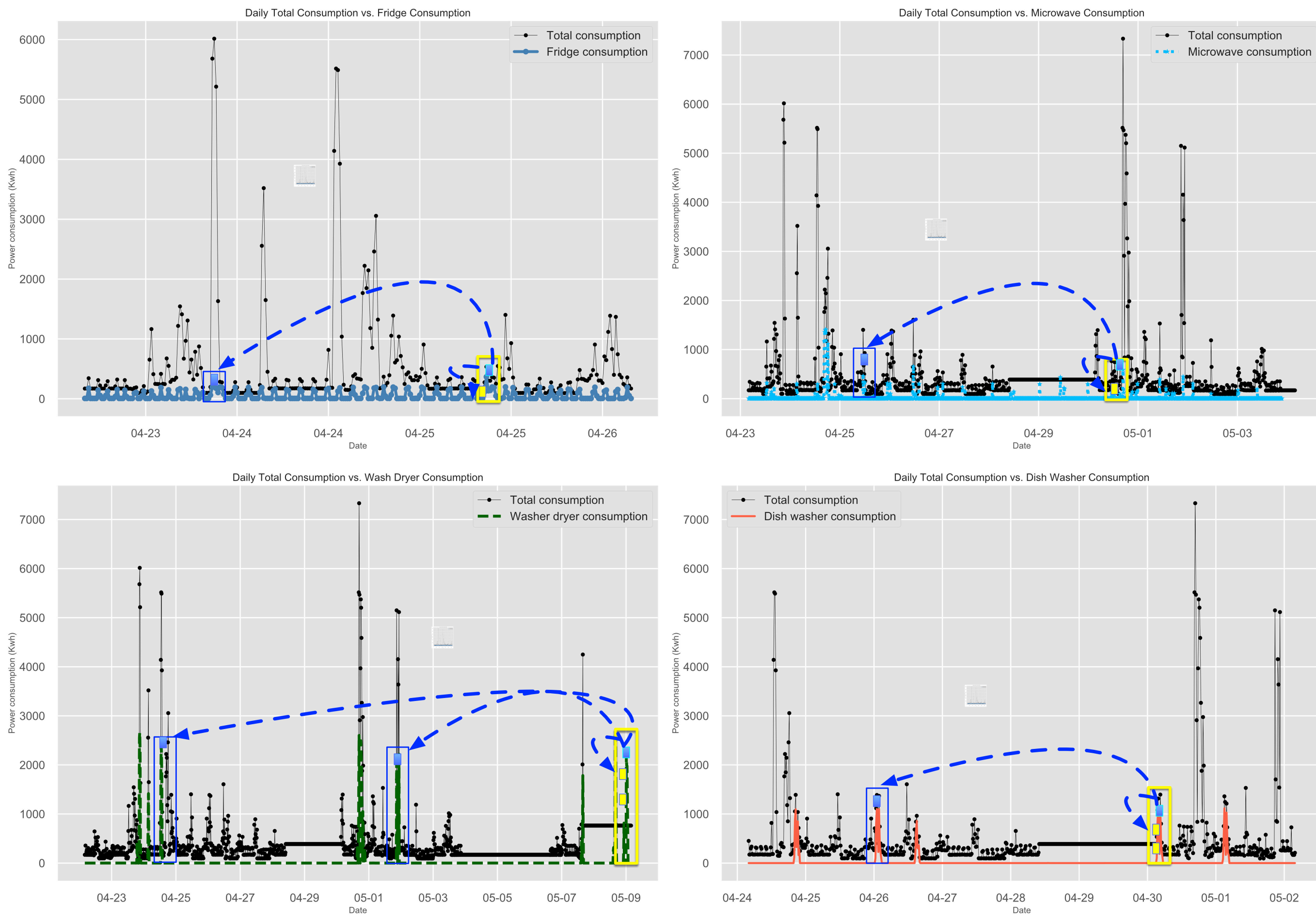

Fig. 1. Different appliances demonstrate various dependencies on short-term, mid-term and long-term neighbors. Appliance-wise consumption patterns in REDD dataset [7] showing temporal dependencies over mixed time scales: The power consumption marked with cyan indicates that it depends on its historical neighbors in a local time window (the yellow rectangle); A potential strong connection with far-apart neighbors (marked in blue points within a blue rectangle window) is also indicated. Depending upon the appliance, such a dependency can range one day ahead (fridge), or even six days ahead (microwave).

Therefore, the ability to use both long-term and short-term dependencies, while varying the attention on them according to the appliance, is crucial in NILM methodologies. How to accomplish this, therefore, is the key question in our paper.

II. Proposed Algorithm and Results

To the best of our knowledge, the adaptive attention is still missing in the current literature and the question of how to put varied attention on different appliances, i.e., short-term or long-term dependency, has not been fully addressed. We therefore propose an Attention-guided Temporal Convolutional Network (ATCN) to encode such dependencies.

Our algorithm sequentially acquires the input and actively attends relevant pieces of temporal information to refine the target consumption estimate at each time step. The key components are the *casual dilation* nature of the model and the *attention mechanisms*, both of which we empirically show the contribution to the appliance-wise consumption prediction. The overall architecture is show in Fig 3.

The comparative results with state-of-the-art algorithms on RMSE metric are shown in Table. I. Our ATCN algorithm has a competitive performance on multi-stage appliances, such as microwave and dish washer (on dish washer, our algorithm can achieve the second best place, just following after Seq2Point algorithm); it particularly achieves the best performance on the most difficult multi-stage appliance: washer dryer, owning to the combination of short-term and long-term dependency and attention mechanism.

In addition to RMSE metric evaluation, we also have provided the comparative results on EA metric, where estimated accuracy (EA) has also been employed as in [8] and [9]. Our approach has proven competitive performance on multi-stage appliances, including wash dryer and dish washer. Be noted that different evaluation metrics does not always have consensus result on different algorithms, which have been observed through the experiments. Moreover, failure to penalize the false detection during the evaluation procedure also has been noticed, and therefore how to propose a reliable evaluation criteria for NILM algorithms would be one of our

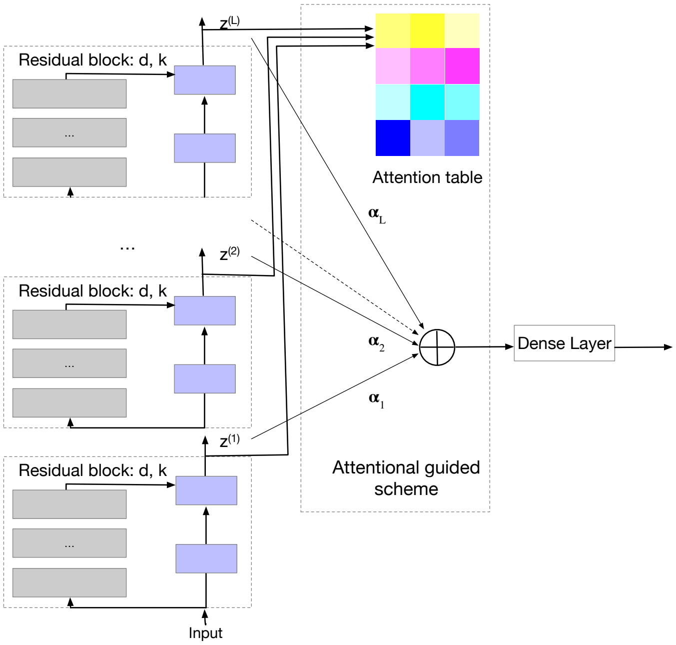

Fig. 2. Attention guided temporal convolutional networks.

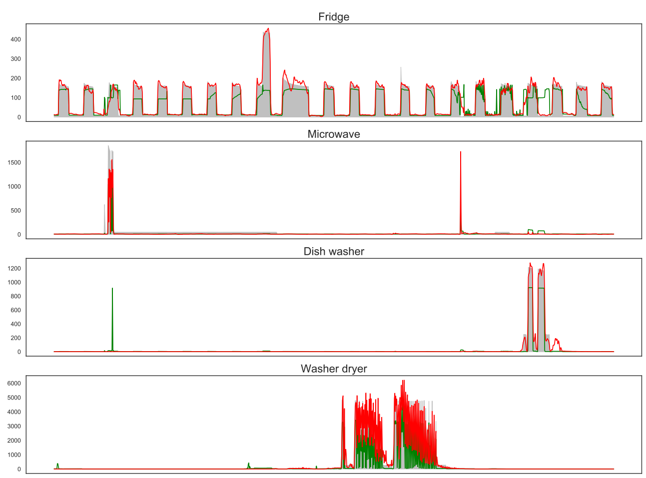

Fig. 3. Comparison of ATCN (red) with TCN (green) algorithm per each appliance on the REDD dataset (replace the last appliance.

research work in future.

III. Conclusion

As an important problem in smart home management, NILM still remains a challenge with great potential for further exploration and improvement. We propose a residual block concatenation strategy and apply an attention mechanism based on such residuals instead of dilated layers to improve NILM performance. The essential dilation and temporal convolution structure helps capture the long-term as well as short-term dependencies in the consumption signatures, while attention residuals ensure that the model's emphasis on relevant time scales is adapted to the appliance. Our proposed

TABLE I
Comparison of RMSE for various models.

Models	RMSE			
	Fridge	Microwave	Dish Washer	Wash Dryer
RNN	16.18	30.58	9.01	316.85
DAE	66.77	52.86	14.83	293.20
Seq2Seq	78.72	33.91	10.65	267.87
Seq2Point	72.52	28.56	4.70	196.10
TCN_NILM	34.71	37.75	40.16	221.82
ATCN_NILM	73.24	44.00	6.59	182.30

TABLE II
Comparison of EA for various models.

Models	EA			
	Fridge	Microwave	Dish Washer	Wash Dryer
RNN	0.97	0.73	0.86	0.50
DAE	0.78	0.39	0.77	0.55
Seq2Seq	0.74	0.69	0.86	0.83
Seq2Point	0.76	0.74	0.94	0.82
TCN_NILM	0.91	0.59	0.29	0.80
ATCN_NILM	0.75	0.55	0.90	0.85

ATCN algorithm outperforms state-of-the-art methodologies on multi-status appliances, especially those with short usage time, and has demonstrated excellent generalization capability.

Acknowledgment

This project was partly supported by Equinor under the Academia Agreement with UiT The Arctic University of Norway, and partly by Kristiania University College under interdisciplinary research fund under the project number: 10096 ForskM 2022- MLG-DA-500T.

References

[1] J. Z. Kolter and T. Jaakkola. "Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation," Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, La Palma, Canary Islands, pp. 1472–1482, 2012.

[2] W. Kong et al. "A Hierarchical Hidden Markov Model Framework for Home Appliance Modeling," in IEEE Transactions on Smart Grid, vol. 9, no. 4, pp. 3079-3090, July 2018.

[3] J. Kelly, and W. Knottenbelt. "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 (BuildSys '15). Association for Computing Machinery, New York, NY, USA, pp. 55–64, 2015.

[4] 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, July 2020.

[5] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton. "Sequence-to-Point Learning With Neural Networks for Non-Intrusive Load Monitoring," AAAI Press, 2019.

[6] Z. Jia, L. Yang, Z. Zhang, H. Liu, and F. Kong, "Sequence to point learning based on bidirectional dilated residual network for non-intrusive load monitoring," International Journal of Electrical Power & Energy Systems, Volume 129, 2021.

[7] J. Z. Kolter, M. J. Johnson, "Redd: A public data set for energy disag- gregation research," in: IN SUSTKDD, 2011.

[8] H. Ren, F. Bianchi, J. Li, R. Olsen, R. Jense, S. Anfinsen, Towards applicability: A comparative study on non-intrusive load monitoring algorithms, in: ICCE, IEEE, United States, 2021.
[9] C. Klemenjak, S. Makonin, W. Elmenreich, Towards comparability in non-intrusive load monitoring: On data and performance evaluation, in: ISGT, USA, IEEE, 2020, pp. 1–5.