2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on

Improving Load Forecast Accuracy of Households Using Load Disaggregation Techniques

Aida Mehdipour Pirbazari, Mina Farmanbar, Antorweep Chakravorty, Chunming Rong
Department of Electrical and Computer Engineering, University of Stavanger, Stavanger, Norway

***Abstract*—The growing use of smart meters in residential buildings provides access to large volumes of consumer data with low-sampling frequency for demand-side analysis. Unlike system-level aggregated data, the measurements from smart meters show more stochastic and irregular behaviour. Advanced load forecasting techniques to deal with irregularities in household load profiles are therefore crucial to develop sustainable and reliable smart grids at local levels. In this paper, we propose a hybrid approach enabling the use of high-resolution smart meter data for short-term load forecasting by incorporating Non-Intrusive Load Monitoring (NILM) technique as a pre-processing step. In particular, the input to the forecasting model includes not only the aggregated energy consumption but also the appliance-level data provided by energy disaggregation algorithms. The experiments on real data sets known as UKDALE and REFIT demonstrate significant improvements in the performance of two forecasting algorithms based on CART and ANN methodologies powered by the NILM input.**

***Keywords*—non-intrusive load monitoring, load disaggregation, short-term load forecasting, MLP, GBRT, FHMM, DAE, Sequence2Sequence**

I. Introduction

Residential consumers make up a significant proportion of end-use electricity demand among different consumer groups. The U.S. Energy Information Administration missions that around 30% of global energy demand will be allocated to the residential sector by the end of 2020 [1]. Households consume energy for heating, cooling, cooking, lighting, and small appliances. The household's demand-side response (DSR) programs provide a positive impact on the power system stability by considering such loads as the active participation of the grid system. The advent of smart grids has contributed to innovative business models and applications, including the short-term load forecasting (STLF) at a local scale which is essential for residential DSR program. Prediction of peak and off-peak periods with short horizons facilitate DSR with running flexible pricing schemes for the households to use most of their electricity during off-peak, and thus reducing their electricity costs.

The techniques aiming to predict the energy consumption of the buildings, as discussed in [2] are classified in two broad categories; engineering (physical) and statistical. The techniques in the first category apply mathematical equations to present the thermal performance of the systems and components of the buildings. While highly accurate and reliable, they are difficult to generalize and require a high degree of detail and expertise to perform costly and elaborate computations. The second category, on the other hand, does not need such detailed data about the simulated building and instead learns from real-time or historical data. The techniques in this category are further categorized into two subgroups: parametric and non-parametric. The parametric methods utilize the past consumption data as well as the principals of regression and time series techniques to perform load forecasting. For example, the studies focusing on Multiple Linear Regression (MLR) [3], Auto-Regressive Integrated Moving Average (ARIMA) [4], [5] and Seasonal ARIMA (SARIMA) [6] fall in to this category. The non-parametric methods, however, focus on load-related and non-related variables as well as various training algorithms to perform short-term load forecasting. The studies which employ machine learning and Artificial-Intelligence-based techniques such as SVR [7], CART [8], ANN [9] and neuro-fuzzy systems [10] belong to this subgroup.

Despite several efforts done by researchers, the task of STLF specifically at a lower aggregation level such as a household level remains challenging due to uncertainty and volatility in load consumption profiles. The residential loads are usually affected with different degrees by several fluctuating factors such as outdoor and indoor temperature, used appliances, installed equipment and the consumption behaviour of the households [11].

Based on the current state-of-the-art, the techniques to deal with volatility at local levels aiming at improving load forecasting can be divided into three categories: (i) Clustering and classification techniques where the customer profiles are clustered based on distinguishing variables such as geographical place, economic parameters as well as consumption load patterns [12], [13]. This approach avoids uncertainty by aggregating the data at higher levels such as a community, a zone or a city. (ii) Methods that use signal processing techniques such as Wavelet Transform [14] and Empirical Mode Decomposition [15] to separate the regular patterns from anomalies, noise and uncertainty of load profile. (iii) Data-driven techniques recently based on deep-learning [16], [17], [18] and [19] which utilize a large amount of aggregated consumption data with high granularity, to enhance load forecasting in terms of accuracy and efficiency.

In recent years, the number of studies based on deep-learning is on the increase due to the roll-out of smart meters and the increasing availability of data with high sampling rates. This growing amount of raw data has also enabled Non-Intrusive Load Monitoring (NILM) with minimal maintenance and installation costs. NILM refers to a process where household total electrical load measured at a single point is decomposed into individual end-appliances signals [20]. There have been several types of research on NILM to provide appliance-level feedback which are discussed in detail in Section II. However, only a few studies, such as [21] and [22], incorporated the NILM to load forecasting problems.

978-1-7281-7647-5/20/$31.00 ©2020 IEEE

DOI 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics50389.2020.00140

In this paper, we investigate the applicability of disaggregation algorithms for load forecasting problem using machine learning and deep learning techniques. For this purpose, a hybrid approach is presented to predict the hourly consumption of households using NILM as a preprocessing step. In particular, the information of energy usage of appliances as a result of the energy disaggregation algorithm is added to the aggregated energy consumption data to provide the forecasting model with more predictive features. The presented model is expected to successfully deal with the uncertainty of smart-meter data and improves household level short-term load forecasting. To test the hybrid method, different combinations of three disaggregation algorithms with two standard forecasting algorithms are trained on real data from two NILM datasets.

The remaining four sections of the paper are organized as follows: Section II, which is related to energy disaggregation problem, consists of five subsections: Subsection A gives an overview of NILM algorithms in the area of household load monitoring; Subsection B briefly introduces the applied NILM algorithms; Subsections C and D explain the disaggregation experiments as well as evaluation metrics. Subsection E provides the disaggregation results besides the performance analysis. Section III, which is related to the forecasting problem, includes three subsections: Subsections A and B describe the methodology of forecasting and error metrics followed by Subsection C presenting the forecasting results. Section IV discusses and compares our findings with the ones in similar related work. Section V, in the end, summarises and concludes our study.

II. Load disaggregation (NILM)

A. States-of-the-arts NILM Algorithms

The learning algorithms which have been employed in the literature of NILM are either supervised or unsupervised. In the supervised approach, both the aggregated data and sub-metered appliances data (as labelled observations) should be collected from the target building to be used in the training phase of the NILM algorithm. However, in terms of metering requirements, the supervised approach can be expensive and time-consuming and difficult to generalize due to its reliance on the building data under study. Some existing works which used SVM and K-means Clustering techniques in NILM [23] belong to this category. The applied models in the unsupervised approach, in contrast, do not need the submetered signals. They only require the aggregate load of the target building to learn and disaggregate loads of appliances. Current NILM research focuses on unsupervised method, which is less expensive and benefits from higher generalization ability.

The methods using in unsupervised approach are further classified into three subgroups as suggested by [24]. The models in the first category such as the ones based on HMM and its variants [25], [26] build the appliance model either manually or during training phase using unlabeled training data. The models in the second group conversely, are trained and developed using labelled data from a known building to be applied later for unknown buildings. Most studies in NILM based on artificial neural networks and deep learning models such as CNN [27], [28] Auto Encoder and LSTM [20], [29], [30] focus on this approach. The last group are fully unsupervised techniques which do not require any prior knowledge of the building or sub-metered data. Their training takes place at the same time as energy disaggregation occurs. The studies such as graph signal processing [31] fall into this category [32].

B. Appplied NILM Algorithms

In this study, we apply three disaggregation algorithms from unsupervised category: (i) 'FHMM' from the first subgroup, (ii) 'Denoising Auto Encoder (DAE) 'and (iii) 'Sequence to Sequence (S2S)' from the second subgroup. All the algorithms are adopted from an open-source toolkit for NILM known as NILMTK [33]. NILMTK is a Python-based open-source framework designed to allow comparison of NILM algorithms across various data sets. It includes several dataset parsers, statistics for dataset analysis, preprocessing and disaggregation algorithms as well as a range of evaluation metrics. We chose the three disaggregation algorithms based on their availability in the MILMTK toolkit as well as the diversity in their architectures and decomposition techniques.

1) Exact Factorial Hidden Markov Model (FHMM): An FHMM is an extension of the Hidden Markov Model (HMM) with hidden states whose observations are modelled by a probability distribution function. In FHMM, there are multiple independent hidden states sequences, and each observation is dependent on several latent variables. Each appliance in FHMM is represented by a hidden Markov model including states and the transition between them. For each state, the average power demand is calculated, and the probabilities of transitions are extracted. Then the aggregate power observations are assigned to the devices whose models require matching with the shape of the appliance signature. More details about this technique can be found in [26].

2) Denoising Auto Encoder (DAE): Autoencoders are a specific type of feedforward neural networks which use sparse intermediate representations to reconstruct their original input at their output. The Denoising Autoencoder (DAE) instead of merely copying their input data, learn to remove noise from the original input. In the NILM context, the total consumption signal is considered as a noisy representation of the target appliance signal, and the DAE eliminates the contribution of other appliances from the aggregate load signal. The adaptation of the DAE method for energy disaggregation problem has been thoroughly explained in [20].

3) Sequence-to-Sequence Optimization (S2S): An S2S architecture uses deep neural networks where the input sequence as the aggregate power readings is mapped to an output sequence as the target appliance power. The input and output sequences are framed as sliding windows such that only a subset of windows can be used for training, thus leading to a reduction in the computational cost. This mapping would help the network to match the seen patterns in the main signal with the features of the appliance signal during training. The S2S approach is defined more extensively in [34].

Implementation Settings: The implementation and default parameter settings of the disaggregation algorithms were taken from the NILMTK-contrib repository[1]. The Denoising Auto

[1] Available at https://github.com/nilmtk/nilmtk-contrib

Fig. 1: Sample disaggregation plots per method for the target appliances (Case study 1)

Encoder and Sequence to Sequence algorithms were trained with a batch size of 1024 with time intervals of 60 s and for a duration of 40 epochs.

C. Disaggregation Experiments

We run the disaggregation algorithms on data from two widely used datasets; REFIT (from Scotland houses) [35] and UKDALE (from England houses) [36]. These datasets contain both aggregate consumption data of the whole house and sub-metered readings from the individual electrical appliances. They were selected, as they include extended periods of historical data (ranging between six months to four years), high sampling rates (ranging between 6 and 8 seconds) and common types of appliances. The first two characteristics are suitable for training the disaggregation and forecasting algorithms. At the same time, the third one serves our purpose to perform a reasonable comparison between the load forecast accuracy of the house profiles with the same type of disaggregated signals.

To investigate the performance of three algorithms in various situations, we present three case studies. In the first case, we train and test the performance of the disaggregation algorithms on one house from REFIT dataset, where one year of the input data is used for training and the remaining three months are used for testing. This experiment would show us how much the disaggregated algorithms can correctly model appliances of a house which was observed by the models during the training. In the second and third cases, we test the performance of the trained models on unseen houses; one house from the same dataset (REFIT) and one house from another dataset (UKDALE) respectively. These experiments would help us to assess the generalization ability of the models within and across the datasets on a small scale.

To provide a fair comparison, we fix the test period in all cases to three months. Besides, the same type of devices across the test houses are chosen for investigation. We select five types of appliances from the two datasets. Mainly; we use (i) washing machine (WM), dishwasher (DW), tumble dryer (TD), kettle (KT) and microwave (MW) of House 5 in REFIT; (ii) WM, DW, TD and KT of House 7 in REFIT and (iii) WM, DW, KT and MW of House 2 in UKDALE. The choice of appliances is driven by the fact that all houses use them, and at least three of selected devices have a high share of daily energy usage and contribute more than 8 per cent to the aggregate load.

D. *Evaluation Metrics*

The metrics which are commonly used for the evaluation of energy disaggregation task, are utilized here: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Normalized RMSE (NRMSE). These metrics measure the deviations between the appliance loads (ground truth) and the disaggregated data for given appliances. Therefore, low error values indicate good disaggregation accuracy. They are defined as follows:

$$MAE = \frac{\sum_{t=1}^{T} |\hat{x}_j(t) - x_j(t)|}{T} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T} (\hat{x}_j(t) - x_j(t))^2}{T}} \tag{2}$$

$$NRMSE = \frac{RMSE}{x_j(max) - x_j(min)} \tag{3}$$

Where $x_j(t)$ denotes appliance j actual power at time step t, $\hat{x}_j(t)$ denotes the estimated power for the jth device at the tth time step, and $x_j(max)$ and $x_j(min)$ denote maximum and minimum power of appliance j during the total observed time-frame T.

Moreover, as recommended in [37] to enhance comparability in NILM on data and performance evaluation, we adopt and add two more metrics to our evaluations: 'noise-to-aggregate-ratio' (NAR) to assess the noise level in aggregate power signal of each house data and 'event ratio' (EVR) to quantify the number of events of target appliance in the test set. They are defined as follows:

$$NAR = \frac{\sum_{t=1}^{T} |y_t - \sum_{j=1}^{M} x_j(t)|}{\sum_{t=1}^{T} y_t} \tag{4}$$

$$EVR = \frac{Events\, of\, target\, appliance\, in\, testset}{Events\, of\, target\, appliance\, in\, dataset} \tag{5}$$

Where in Eq.4 M is the total number of sub-metered signals (appliances), y_t is the aggregate power signal and $x_j(t)$ is the power consumption of appliance j at time t and in Eq.5 the events refer to the activation event of the appliance (The number of times when the appliance is On).

E. *Disaggregation Results and Analysis*

Fig. 1 shows a sample of temporal behaviour performance of the three disaggregation algorithms in the first case study. As we can see, DAE and Seq2Seq models perform better than FHMM in detecting the time intervals when the target device consumes electricity. FHMM is also weak at the correct estimation of power consumption of the device during 'off' and 'on' periods. For example, even if the dishwasher is not consuming energy, it overestimates the appliance consumption, and when it starts working, the energy usage is underestimated by FHMM. Moreover, the Seq2Seq model, compared to the other two, is more successful at following the consumption pattern of the appliance. For example, based on the plots, the consumption pattern of the washing machine, kettle and microwave are detected with higher accuracy using the Seq2Seq algorithm.

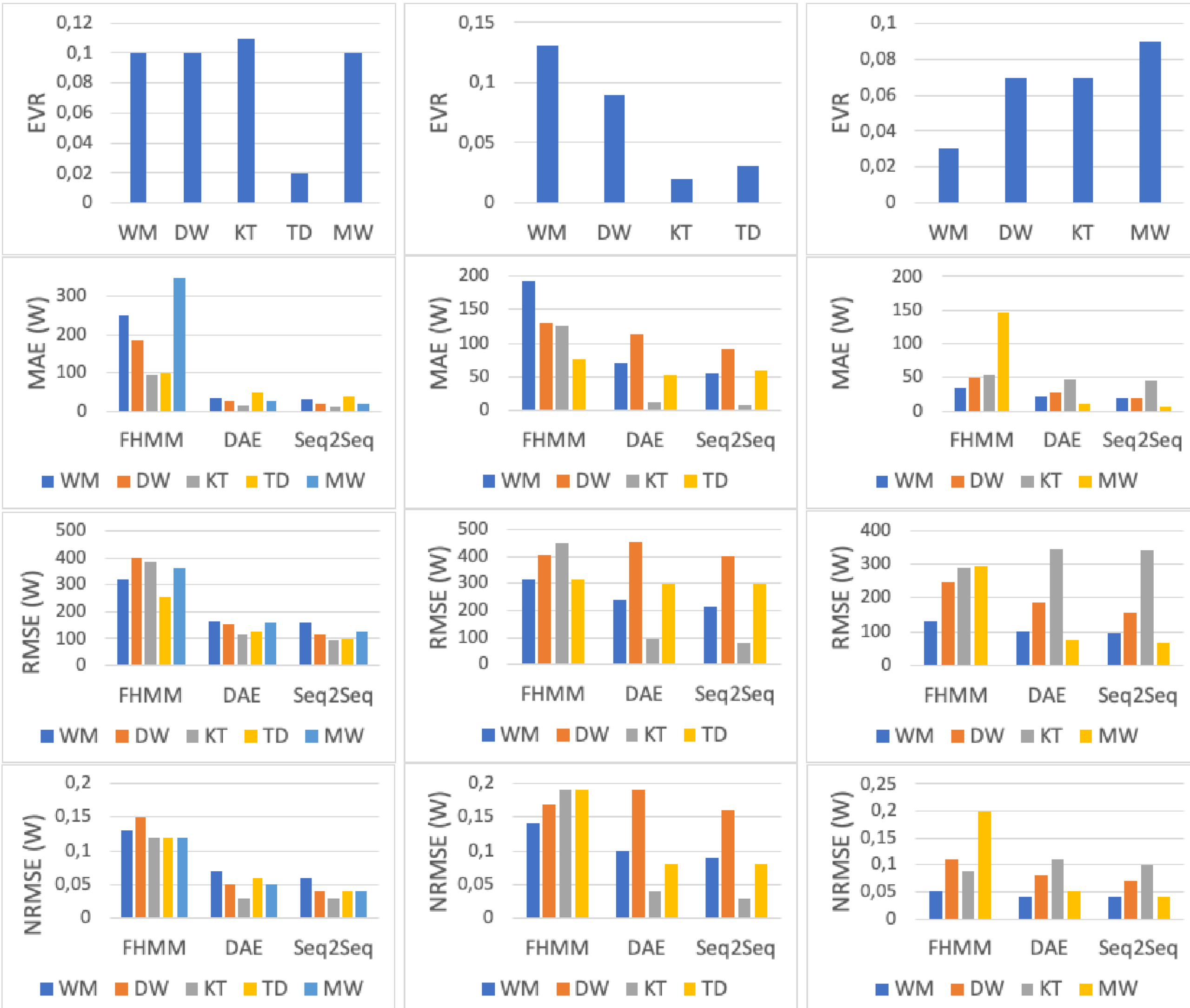
Fig .2 presents the comparative results for all case studies in three columns. The first bar chart in each column shows the EVR value per appliance for the corresponding cases study. The next three diagrams in each column illustrate the performance (MAE, RMSE, NRMSE) per method for the selected devices. As shown, on both seen house (Case study 1) and unseen houses (Case study 2 and 3) during training, the Seq2Seq and DAE outperform FHMM on most appliances on every metric. For example, these models in Case 1, compared to FHMM, have experienced around 80 per cent reduction in MAE values on average for five target appliances. In Case 2, however, they are not as successful as in the other cases showing significantly higher error values specifically for multi-state devices (washing machine, dishwasher and tumble dryer). Regarding the EVR value, it is expected that the test set with fewer events yields a lower disaggregation error, and therefore, better evaluation results. In our experiments, this expectation comes true most for the FHMM model where the smaller EVR values of all tested appliances (except for the kettle) in Case 3 has led to a considerable reduction in MAE and RMSE values.

Regarding the NAR metric, we should mention that House 7 and House 5, compared to House 2 with larger NAR values show higher levels of noise in their aggregate power signals (0.61 and 0.51 versus 0.35). Generally, this can make the disaggregation more challenging if the contributing appliances in the noisy datasets would not play an essential role in the total consumption. In our study, although high-consuming appliances such as tumble dryer and washing machine are selected from noisy datasets, yet the complexity in the disaggregation task is visible. For example, all three algorithms in House 7 (Case 2) experience high RMSE values of more than 200 watts for the majority of appliances.

III. Short-term Load Forecasting

A. *Forecasting Methods and Model Development*

In this work, we choose two machine learning techniques, namely Multilayer Perceptron (MLP) and Gradient Boosting Regression Tree (GBRT) for short-term load forecasting. The first one belongs to the category of ANN-based models and along with its variants is one of the most applied techniques in the STLF problems. The second one with combining the advantages of decision tree algorithms and ensemble technique has shown promising results in different regression tasks,



(a) Case study 1 (b) Case study 2 (c) Case study 3

Fig. 2: Disaggregation results per method for the target appliances per case study

including short-term load forecasting. More details on the architecture and application of MLP and GBRT techniques can be found in [38] and [39] respectively.

As mentioned in the introduction section, to build a hybrid model, we enrich the inputs with information from the studied appliances. Particularly, the input vector to both MLP and GBRT models includes two types of features. The first type consists of the variables representing hourly power demand of the home appliances which are extracted from the energy disaggregation step (washing machine, dishwasher, kettle, microwave or tumble dryer). The second type refers to five variables relating to the total consumption of the house including power demand of the current hour, previous two, three, 23 and 24 hours and average power demand of the present day.

To improve the accuracy of forecasting, we adopt Rolling Update (RU) method to use the known information of the current day. This method dynamically updates the input of the model in real-time and thus provides useful information for forecasting the future values of load demand. The main idea of RU is that the input data (load value at time t) after each time step is supplemented by new data (observed load value at time t+1) in a rolling way. Afterwards, the old data (the most past load value) is removed so that the length of the data remains unchanged. Therefore, the created input vectors will better represent the current features of the prediction system [40]. Each forecasting model has one output which represents the estimated total consumption of the house during the next hour of the day. For compatibility, all the measurements from each home are downsampled to one-hour resolution.

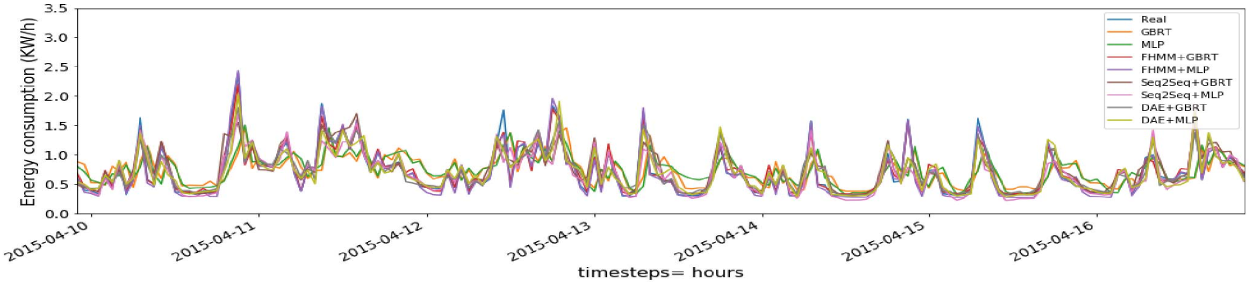


Fig. 3: Hourly predictions versus real consumption of House 5 (REFIT) over one week

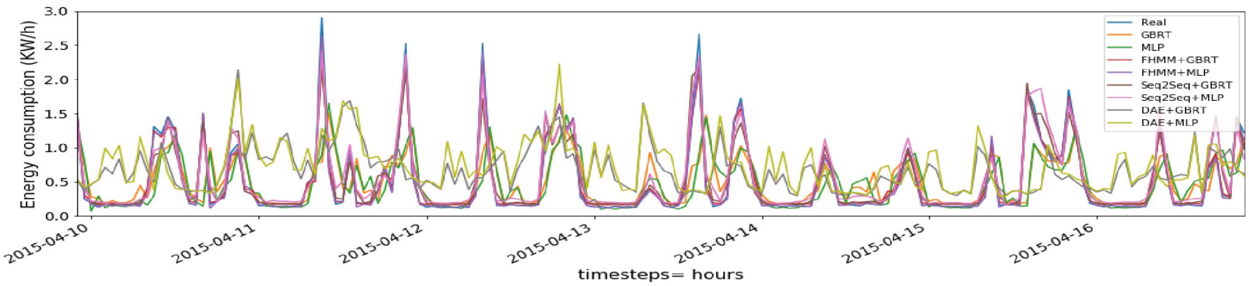


Fig. 4: Hourly predictions versus real consumption of House 7 (REFIT) over one week

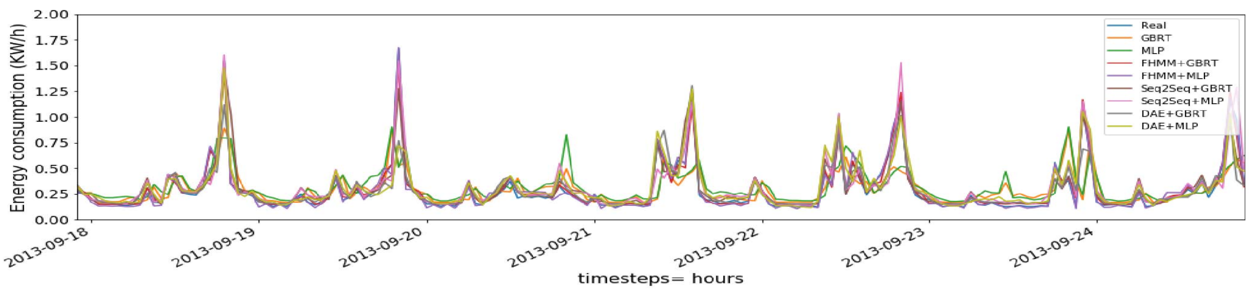


Fig. 5: Hourly predictions versus real consumption of House 2 (UKDALE) over one week

Implementation Setup: The models were developed in Python using scikit-learn library. Regarding the MLP, we considered a single hidden layer with n input nodes corresponding to the number of input variables (9 inputs for House 7 and House 2 with four appliances; 10 inputs for House 1 with five devices). The number of nodes in the hidden layer was set as twice as the number of input variables plus one. 'Adam' was used as the optimization algorithm along with 'Uniform' weight initialization technique. The Rectified Linear Units (ReLU) was applied in the hidden layer, while the linear activation function is used in the output layer. For the training process, a batch size of 32 training records and 60 epochs were set for each MLP model. Regarding the GBRT, we set the number of trees (estimators) as 300 with a maximum depth of 3. The rest of the parameters were set according to the default values in the scikit-learn package with the version number of 0.23.1.

B. Evaluation Metrics

To evaluate the performance of the hybrid approach, we use the same metrics (MAE and RMSE) as defined in Section II part D, with replacing $x_j(t)$ and $\hat{x}_j(t)$ by the actual and predicted aggregate load of the household. Moreover, we add two specific metrics to load forecasting problems known as Daily peak Mean Average Percentage Error (DpMAPE), adopted from [41] and Daily total Mean Average Percentage Error (DtMAPE). The DpMAPE measures the accuracy of the model in prediction of daily peak consumption and the DtMAPE evaluates the model accuracy in estimation of total daily load consumption. They are defined as follows:

$$DpMAPE = |\frac{y_{max} - \hat{y}_{max}}{y_{max}}| * 100 \tag{6}$$

$$DtMAPE = |\frac{y_{total} - \hat{y}_{total}}{y_{total}}| * 100 \tag{7}$$

Here, y_{max} and $\hat{y}_{max}$ denote the observed and predicted daily peak consumption, and similarly y_{total} and $\hat{y}_{total}$ denote the observed and predicted daily total consumption values.

C. Forecasting Results and Comparison

To demonstrate the predictive superiority of the proposed hybrid approach, we selected the MLP and GBRT models

Model	MAE (KW/h)	RMSE (KW/h)
MLP	0.21	0.30
FHMM+MLP	0.04	0.06
DAE+MLP	0.10	0.19
S2S+MLP	0.12	0.23
GBRT	0.20	0.29
FHMM+GBRT	**0.03**	**0.06**
DAE+GBRT	0.10	0.18
S2S+GBRT	0.12	0.23

(a) House 5

Model	MAE (KW/h)	RMSE (KW/h)
MLP	0.22	0.35
FHMM+MLP	**0.02**	**0.04**
DAE+MLP	0.14	0.23
S2S+MLP	0.09	0.17
GBRT	0.22	0.36
FHMM+GBRT	0.04	0.05
DAE+GBRT	0.12	0.21
S2S+GBRT	0.09	0.16

(b) House 7

Model	MAE (KW/h)	RMSE (KW/h)
MLP	0.10	0.20
FHMM+MLP	**0.01**	**0.02**
DAE+MLP	0.05	0.10
S2S+MLP	0.05	0.09
GBRT	0.11	0.19
FHMM+GBRT	0.02	0.04
DAE+GBRT	0.05	0.10
S2S+GBRT	0.05	0.08

(c) House 2

TABLE I: Forecasting results of different models per dataset

without disaggregated data for the comparison. For comparability, the test period of all houses was set to three months and the training period was considered as one year, except for House 2 whose measurements were only available for seven months. Fig. 3, Fig. 4 and Fig .5 show the results of hourly load predictions of the target house made by the different approaches over a one-week test period. As shown, in most cases, the methods with the disaggregation step follow the consumption pattern more closely compared to the techniques without appliance data (GBRT and MLP). Table. I summarise the forecasting results over the whole test period (three months) for each test house.

According to Table. I, in all experiments, the performance of both forecasting algorithms (MLP and GBRT) has significantly improved by supplementing their inputs with appliance load information. More precisely, for the seen house during the energy disaggregation step (House 5), the hybrid MLP-based approaches, on average brought 56% reduction in MAE and 46% reduction in RMSE compared with MLP. Similarly, the GBRT models with disaggregation inputs on average reached to reduce rates of 58 % in MAE and 48 % in RMSE compared with GBRT. The proposed approaches also show good generalization ability while testing on unseen houses. For House 7, FHMM+MLP outperform MLP with 90% and 88% reductions in MAE and RMSE respectively, and for House 2 the same approach improves the accuracy of MLP up to 90% on both error metrics. Furthermore, the combinations of GBRT with FHMM and Seq2Seq algorithms show similar improvements in the performance of GBRT for both unseen houses. These results can also be concluded from Fig. 6 and Fig. 7, where the daily peak and total MAPE values are reduced on average by 16 % and 30 % respectively utilizing the proposed approaches.

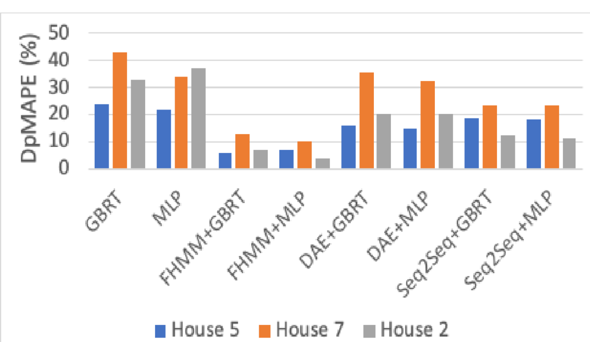


Fig. 6: Daily peak MAPE values by method and test house

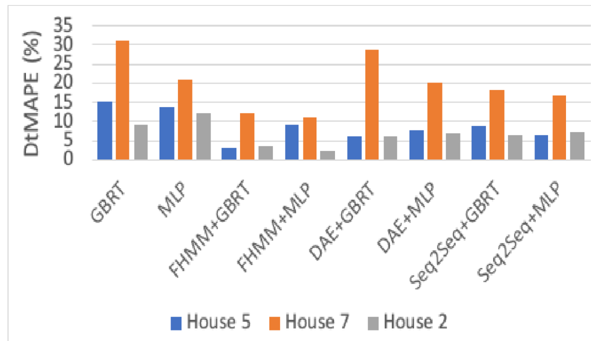


Fig. 7: Daily total MAPE values by method and test house

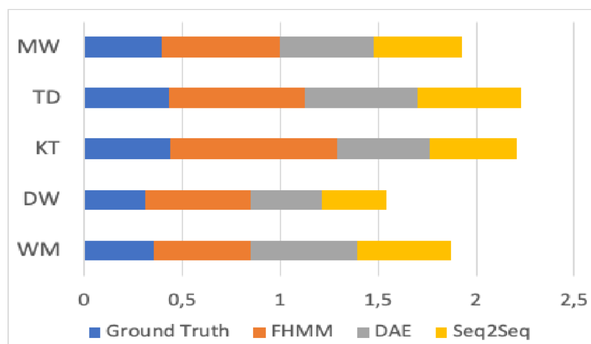


Fig. 8: Correlation values between total consumption and power consumption of appliances in House 5

Regarding the forecasting algorithms, the MLP algorithm has slightly performed better than the GBRT method in load prediction of House 5 and House 7. In contrast, for House 2, GBRT performs load prediction with higher accuracy. Furthermore, among the test houses, the future demand of house 7 seems more challenging to predict, since, in most test scenarios of this house, different approaches have produced significantly higher daily MAPE values (Fig. 6 and Fig. 7).

As for the disaggregation algorithms, Seq2Seq-based models slightly outperform the DAE-based ones. However, the lowest error rates indicating the highest prediction accuracy, are reported for FHMM-based approaches (Table. I). In the first place, this may seem a contradicting result as the FHMM model has not performed as accurately as the other disaggregation algorithms in Section II. However, by further investigation on the disaggregated values and errors produced by this approach, we found out FHMM gains this capability by magnifying the contribution of appliances to the aggregate power signal and therefore falsely feed the forecasting models with the values closer to the aggregate load. Fig .8 can implicitly prove our justification. This figure reports the Pearson correlation values between hourly energy usage of the appliances (ground truth as well as disaggregated) and total hourly consumption of the sample training house. As seen, the larger values (longer rectangles in the plots) mainly belong to the FHMM which means the model can provide the highly correlated information with future total load consumption, and thus enabling the prediction model to generate forecasting results with high precision.

IV. Comparison With a Similar Study

We found one study [22] among related works that have applied a similar approach to ours to enhance short-term load forecasting. In their research, three baseline methods; Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), Feed Forward Neural network (FFNN) and the variants of FFNN utilizing three NILM algorithms have been evaluated on UKDALE dataset. The NILM algorithms include Denoising Auto-Encoder (DAE), a Long Short-Term Memory network (LSTM) and a network that provides rectangles for the estimated demand by regression the start time, end time and average power demand (nicknamed as RECTANGLES). According to their findings, the combination of DAE with the predictive algorithm (FFNN) results in the greatest decrease in RMSE and MAE, whereas the combinations of FHMM and LSTM (Seq2Seq) with FFNN (MLP) provide better accuracy in our research.

However, a direct comparison between the experimental results of the two studies, since the studies are carried out differently, does not seem fair. To be specific, despite some similarities in terms of dataset, error measurements, forecasting and disaggregation algorithms, there are still variations in training and testing profiles, the test period, selected appliances and model parameters that could have a substantial impact on the outcomes. Therefore, we do not draw a general conclusion which disaggregation techniques work better or which combinations are often more accurate than any other hybrid models. We may infer, however, from two studies that hybrid approaches are more effective and achieve much better performance compared to the single models for majority of the cases in the experiments. Besides, we realize that the degree of effectiveness depends on several variables, such as appliance type, training data, testing period, data resolution, and algorithms for disaggregation and forecasting tasks.

V. Conclusions

In this paper, a hybrid approach was presented for predicting hourly load consumption of households using smart meter data. This approach utilizes the results of load disaggregation algorithms to improve the performance of the forecasting algorithms. Three disaggregation algorithms (FHMM, Denoising AutoEncoder and Sequence to Sequence) from NILMTK framework and two forecasting algorithms (MLP and GBRT) were developed and trained on three house profiles and five appliances from two publicly available NILM datasets (REFIT and UKDALE). The results showed that among the disaggregation techniques, the Sequence to Sequence model was more successful in correctly decomposing the load signature of the majority of appliances in unseen and seen houses during training. Furthermore, regarding the forecasting task, the hybrid models specifically FHMM+MLP and Seq2Seq+MLP mostly gained better prediction accuracy as compared to the models without disaggregation step showing at least 80% and 42% reduction in MAE respectively. The results motivate further investigation for a larger number of houses, appliances and NILM algorithms based on deep networks.

References

[1] M. Senn, "U.S. Energy Information Administration, International Energy Outlook." https://www.eia.gov/forecasts/ieo/pdf/0484.pdf, 2016. accessed May 15, 2020.

[2] H.-x. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012.

[3] D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, "Short-term load forecasting using regression based moving windows with adjustable window-sizes," in *2014 IEEE Industry Application Society Annual Meeting*, pp. 1–8, IEEE, 2014.

[4] C.-M. Lee and C.-N. Ko, "Short-term load forecasting using lifting scheme and arima models," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5902–5911, 2011.

[5] D. Alberg and M. Last, "Short-term load forecasting in smart meters with sliding window-based arima algorithms," *Vietnam Journal of Computer Science*, vol. 5, no. 3-4, pp. 241–249, 2018.

[6] L. J. Soares and M. C. Medeiros, "Modeling and forecasting short-term electricity load: A comparison of methods with an application to brazilian data," *International Journal of Forecasting*, vol. 24, no. 4, pp. 630–644, 2008.

[7] E. Ceperic, V. Ceperic, and A. Baric, "A strategy for short-term load forecasting by support vector regression machines," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4356–4364, 2013.

[8] G. Dudek, "Short-term load forecasting using random forests," in *Intelligent Systems' 2014*, pp. 821–828, Springer, 2015.

[9] A. Baliyan, K. Gaurav, and S. K. Mishra, "A review of short term load forecasting using artificial neural network models," *Procedia Computer Science*, vol. 48, pp. 121–125, 2015.

[10] H. J. Sadaei, F. G. Guimarães, C. J. da Silva, M. H. Lee, and T. Eslami, "Short-term load forecasting method based on fuzzy time series, seasonality and long memory process," *International Journal of Approximate Reasoning*, vol. 83, pp. 196–217, 2017.

[11] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91–99, 2016.

[12] F. Fahiman, S. M. Erfani, S. Rajasegarar, M. Palaniswami, and C. Leckie, "Improving load forecasting based on deep learning and k-shape clustering," in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 4134–4141, 2017.

[13] Y. Wang, Q. Chen, M. Sun, C. Kang, and Q. Xia, "An ensemble forecasting method for the aggregated load with subprofiles," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3906–3908, 2018.

[14] J. Zhang, Y.-M. Wei, D. Li, Z. Tan, and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, vol. 158, pp. 774–781, 2018.

[15] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy load forecasting using empirical mode decomposition and support vector regression," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 549–556, 2013.

[16] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, "An overview and comparative analysis of recurrent neural networks for short term load forecasting," *arXiv preprint arXiv:1705.04378*, 2017.

[17] S. Ryu, J. Noh, and H. Kim, "Deep neural network based demand side short term load forecasting," *Energies*, vol. 10, no. 1, p. 3, 2017.

[18] A. Almalaq and G. Edwards, “A review of deep learning methods applied on load forecasting,” in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 511–516, 2017.

[19] A. Mehdipour Pirbazari, M. Farmanbar, A. Chakravorty, and C. Rong, “Short-Term Load Forecasting Using Smart Meter Data: A Generalization Analysis,” *Processes*, vol. 8, p. 484, apr 2020.

[20] 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*, pp. 55–64, 2015.

[21] S. Welikala, C. Dinesh, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, “Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting,” *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 448–461, 2017.

[22] A. F. Ebrahim and O. A. Mohammed, “Energy disaggregation based deep learning techniques: A pre-processing stage to enhance the household load forecasting,” in *2018 IEEE Industry Applications Society Annual Meeting (IAS)*, 2018 pages=1-8.

[23] H. Altrabalsi, J. Liao, L. Stankovic, and V. Stankovic, “A low-complexity energy disaggregation method: Performance and robustness,” in *2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG)*, pp. 1–8, IEEE, 2014.

[24] B. Zhao, L. Stankovic, and V. Stankovic, “On a training-less solution for non-intrusive appliance load monitoring using graph signal processing,” *IEEE Access*, vol. 4, pp. 1784–1799, 2016.

[25] J. Z. Kolter and T. Jaakkola, “Approximate inference in additive factorial hmms with application to energy disaggregation,” in *Artificial intelligence and statistics*, pp. 1472–1482, 2012.

[26] A. Zoha, A. Gluhak, M. Nati, and M. A. Imran, “Low-power appliance monitoring using factorial hidden markov models,” in *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 527–532, IEEE, 2013.

[27] D. de Paiva Penha and A. R. G. Castro, “Convolutional neural network applied to the identification of residential equipment in non-intrusive load monitoring systems,” in *3rd International Conference on Artificial Intelligence and Applications*, pp. 11–21, 2017.

[28] L. Massidda, M. Marrocu, and S. Manca, “Non-intrusive load disaggregation by convolutional neural network and multilabel classification,” *Applied Sciences*, vol. 10, no. 4, p. 1454, 2020.

[29] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis, and A. Doulamis, “Context aware energy disaggregation using adaptive bidirectional lstm models,” *IEEE Transactions on Smart Grid*, 2020.

[30] D. Murray, L. Stankovic, V. Stankovic, S. Lulic, and S. Sladojevic, “Transferability of neural network approaches for low-rate energy disaggregation,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8330–8334, IEEE, 2019.

[31] K. He, L. Stankovic, J. Liao, and V. Stankovic, “Non-intrusive load disaggregation using graph signal processing,” *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1739–1747, 2016.

[32] A. Faustine, N. H. Mvungi, S. Kaijage, and K. Michael, “A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem,” *arXiv preprint arXiv:1703.00785*, 2017.

[33] N. Batra, R. Kukunuri, A. Pandey, R. Malakar, R. Kumar, O. Krystalakos, M. Zhong, P. Meira, and O. Parson, “Towards reproducible state-of-the-art energy disaggregation,” in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 193–202, 2019.

[34] C. Zhang, M. Zhong, Z. Wang, N. Goddard, and C. Sutton, “Sequence-to-point learning with neural networks for non-intrusive load monitoring,” in *Thirty-second AAAI conference on artificial intelligence*, 2018.

[35] D. Murray, L. Stankovic, and V. Stankovic, “An electrical load measurements dataset of united kingdom households from a two-year longitudinal study,” *Scientific data*, vol. 4, no. 1, pp. 1–12, 2017.

[36] J. Kelly and W. Knottenbelt, “The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes,” *Scientific data*, vol. 2, no. 1, pp. 1–14, 2015.

[37] C. Klemenjak, S. Makonin, and W. Elmenreich, “Towards comparability in non-intrusive load monitoring: On data and performance evaluation,” in *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2020.

[38] G. Dudek, “Multilayer perceptron for short-term load forecasting: from global to local approach,” *Neural Computing and Applications*, pp. 1–13, 2019.

[39] M. D. C. Ruiz-Abellón, A. Gabaldón, and A. Guillamón, “Load forecasting for a campus university using ensemble methods based on regression trees,” *Energies*, vol. 11, no. 8, p. 2038, 2018.

[40] S. Wang, X. Wang, S. Wang, and D. Wang, “Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting,” *International Journal of Electrical Power & Energy Systems*, vol. 109, pp. 470–479, 2019.

[41] B. Asare-Bediako, W. Kling, and P. Ribeiro, “Day-ahead residential load forecasting with artificial neural networks using smart meter data,” in *2013 IEEE Grenoble Conference*, pp. 1–6, IEEE, 2013.