NTUA

BAYESIAN-OPTIMIZED BIDIRECTIONAL LSTM REGRESSION MODEL FOR NON-INTRUSIVE LOAD MONITORING

ICASSP

Maria Kaselimi[(1)], Nikolaos Doulamis[(1)] Anastasios Doulamis[(1)] Athanasios Voulodimos[(1,2)]
Eftychios Protopapadakis[(1)]

[(1)]National Technical University of Athens, GR, [(2)]University of West Attica, GR

mkaselimi@mail.ntua.gr, ndoulam@cs.ntua.gr, adoulam@cs.ntua.gr, thanosv@mail.ntua.gr, eftpro@mail.ntua.gr

National Technical University of Athens
Greece

2019 IEEE International Conference on Acoustics, Speech and Signal Processing
12 – 17 May, 2019 | Brighton, UK.

ABSTRACT

In this paper, a Bayesian-optimized bidirectional Long Short -Term Memory (LSTM) method for energy disaggregation, is introduced. Energy disaggregation or the so called Non-Intrusive Load Monitoring (NILM), is a process aiming to identify the individual contribution of appliances in the aggregate electricity load. The proposed model, called Bayes-BiLSTM is organized in a modular way to address multi-dimensionality issues that arise when the number of appliances increase. In addition, a non-causal model is introduced in order to tackle with inherent structure, characterizing the operation of multi-state appliances. Furthermore, a Bayesian-optimized framework is introduced to select the best configuration of the proposed regression model, thus increasing performance. Experimental results indicate the proposed method's superiority, compared to the current state-of-the-art.

Contribution

The proposed appliance-based, Bayesian-optimized BiLSTM regression model satisfies a set of crucial characteristics making it superior than the other NILM (Non-Intrusive Load Monitoring) methods. These specific features are summarized below:

- **Long Term Regression:** This work addresses NILM as a sequence-to-sequence regression problem, thus allowing to maintain all the necessary information. Additionally, existing long term dependencies should be accounted for, increasing regression performance.
- **Modularity:** Our approach is conducted for each device separately with an appliance-based, modular and extensible model, thus addressing dimentionality issues.
- **Optimization:** Bayesian optimization strengthens model's performance through the optimal hyperparameters selection, creating a unique optimal model, adaptable to each appliance's individual settings and seasonal variations.
- **Non-causality:** In our approach, non-causality is achieved by modifying the conventional LSTM (Long Short-Term Memory) network taking into account both previous and future states of electricity power load. Therefore, bidirectional, recurrent regression deep models are adopted for NILM.

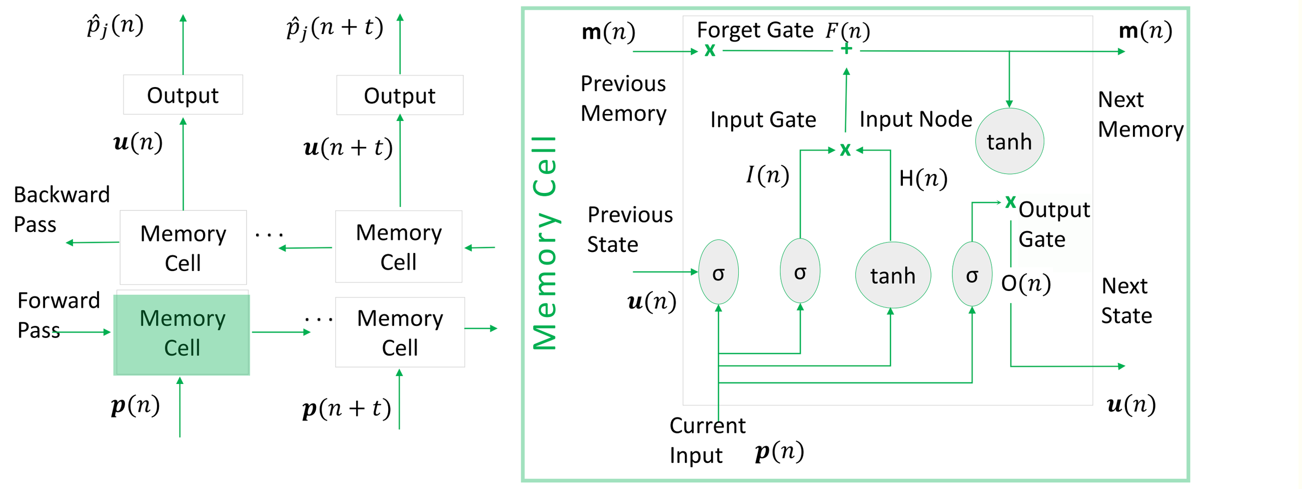


Fig.1 Bidirectional long range recurrent regression model and the respective memory cell.The memory cell of a LSTM network. It contains three different components; (i) the forget gate $F(n)$, (ii) the input gate $I(n)$ and the input node $H(n)$ and (iii) the output gate $O(n)$.

Methodology

Let M be a set of all known household's appliances. Let $p(n)$ be the aggregate measured energy signal at time t. Let us denote by $p_j(n)$ the active power load of j-th appliance out of M available.We can express $p(n)$ as [1]:

$$p(n) = \sum_{j=1}^{M} p_j(n) + e(n)$$

where $e(n)$ is the noise of the measurements. So, the values $p_j(n)$ expressed as:

$$p_j(n) = f(\mathbf{p}(n)) + e(n) = \hat{p}_j(n) + e(n)$$

where $\mathbf{p}(n) = [(p(n) \cdots p(n-K))]^T$ is the aggregate signal $p(n)$ over a time window K+1 and $f(\cdot)$ is a non-linear function. One way to approximate the unknown relationship $f(\cdot)$ is through a feed-forward neural network:

$$\hat{p}_j(n) = [\mathbf{u}_j(n)]^T \cdot \mathbf{v}_j$$

$$\mathbf{u}_j(n) = \begin{bmatrix} u_{j,1}(n) \\ \vdots \\ u_{j,L}(n) \end{bmatrix} = \begin{bmatrix} \tanh(\mathbf{w}_{j,1}(n) \cdot \mathbf{p}(n)) \\ \vdots \\ \tanh(\mathbf{w}_{j,L}(n) \cdot \mathbf{p}(n)) \end{bmatrix}$$

where $\mathbf{w}(j,i) = , i = 1, , L$, are weights connecting the input and the i-th hidden neuron and vector $\mathbf{u}(n)$ is a state vector gathering all hidden layer responses $u(j,i)$ at n time period. These non-linear transformations are linearly combined to provide the estimate of $\hat{p}_j(n)$, using a set of weights $\mathbf{v}_j$.

TABLE I. Performance evaluation has been performed among Bayes-BiLSTM and other approaches, such as CNNs, unidirectional LSTMs, CO and FHMM. The table presents the comparative results based on objective metrics of MAE, RMSE and NRMS. In this experimental setup, we have selected four appliances of dataset, which are CDE, DWE, HPE, WOE.

	MAE	RMSE	NRMS	MAE	RMSE	NRMS
Methods	*Appliance 1: CDE*			*Appliance 2: DWE*		
Bayes-BiLSTM	9.19	62.19	0.15	6.43	51.14	0.27
LSTM	25.35	180.91	0.43	24.04	72.05	1.00
CNN	34.42	202.52	0.48	32.67	84.59	1.24
CO	117.53	323.23	1.26	156.23	317.62	4.41
FHMM	129.57	453.31	0.90	313.68	459.07	4.44
Methods	*Appliance 3: HPE*			*Appliance 4: WOE*		
Bayes-BiLSTM	106.56	395.29	0.59	8.06	118.63	0.75
LSTM	161.74	369.81	0.55	15.82	141.2	0.89
CNN	158.68	305.47	0.46	23.30	138.7	0.88
CO	249.16	426.66	1.23	267.00	360.85	3.46
FHMM	121.69	458.77	1.14	49.38	432.79	2.89

MAE	$\frac{1}{n}\sum_{j=1}^{M}\lvert p_j(n) - \hat{p}_j(n)\rvert$	RMSE	$\sqrt{\frac{1}{n}\sum_{j=1}^{M}(p_j(n) - \hat{p}_j(n))^2}$	NRMS	$\sqrt{\frac{\sum_{j=1}^{M}(p_j(n) - \hat{p}_j(n))^2}{\sum_{j=1}^{M} p_j^2(n)}}$

Since the appliances randomly become dynamically active/inactive, the state vector $u(n)$ depends on its previous values. So,

$$u_i(n) = g(\mathbf{w}_i^T \cdot \mathbf{p}(n) + r_i^T \cdot \mathbf{u}(n-1))$$

where r_i is a set of parameters that weigh the contribution of $u(n-1)$ to the current state values. This equation models the Recurrent Neural Network (RNN). Forming the bidirectional RNN, gives:

$$u_i(n) = g(\mathbf{w}_i^T \cdot \mathbf{p}(n) + \overrightarrow{\mathbf{r}}_i^T \cdot u(n-1) + \overleftarrow{\mathbf{r}}_i^T \cdot \mathbf{u}(n+1))$$

Finally, a bidirectional LSTM network [2] is adopted as the basic regression model for power load estimation.

$$\{f(n), I(n), h(n), O(n)\} = \{\sigma, \tanh\}(\mathbf{w}_i^{T,C} \cdot \mathbf{p}(n) + \overrightarrow{\mathbf{r}}_i^{T,C} \cdot u(n-1) + \overleftarrow{\mathbf{r}}_i^{T,C} \cdot \mathbf{u}(n+1))$$

where $C = \{f, I, h, O\}$.

Experimental Evaluation

In this section, the proposed modular, optimized and con-text-aware model, called Bayes-BiLSTM is evaluated. Specifically, the Bayes-BiLSTM method is compared against other state of the art deep learning methods (i.e., LSTM, CNN), which lack bidirectionality and context adaptivity. We also compare the aforementioned results with the state-of-the-art NILM algorithms i.e., FHMM-based and CO-based methods from NILMTK [3] which are widely used in energy disaggregation research. Table 1 shows the performance metrics for four selected appliances, which are clothes dryer (CDE), dishwasher (DWE), heat pump (HPE) and wall oven appliance (WOE) of **AMPds** dataset [4].

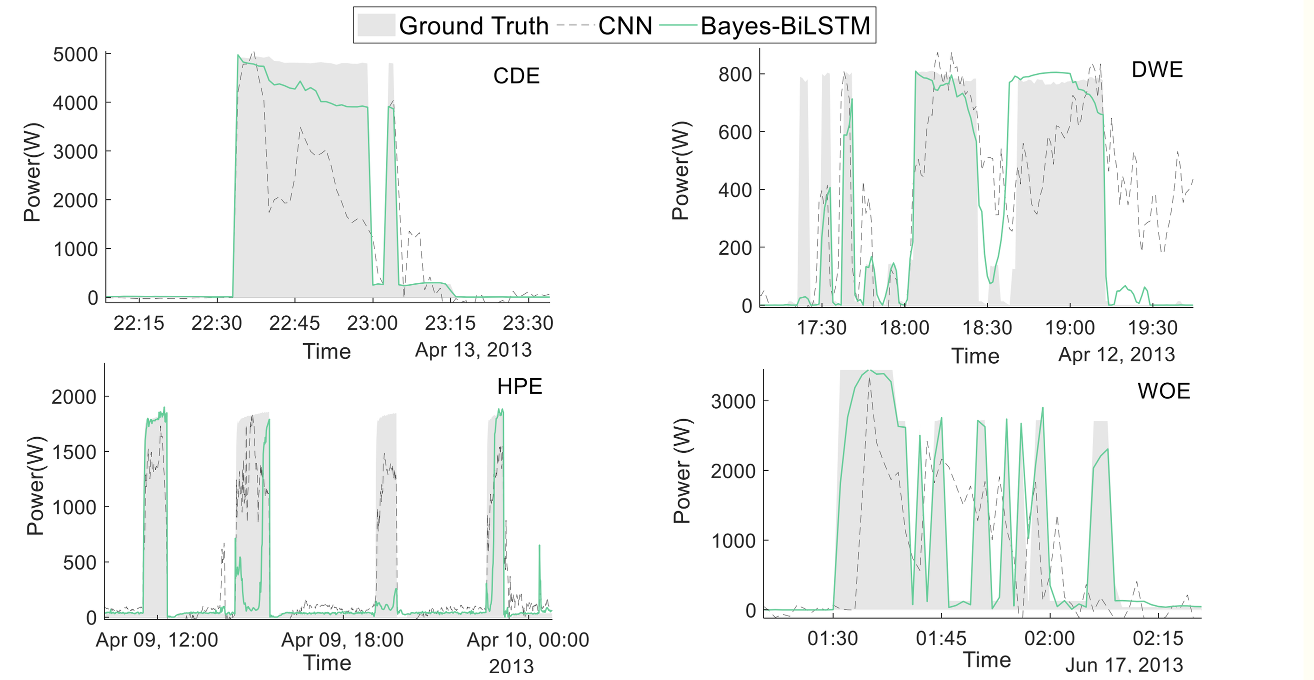
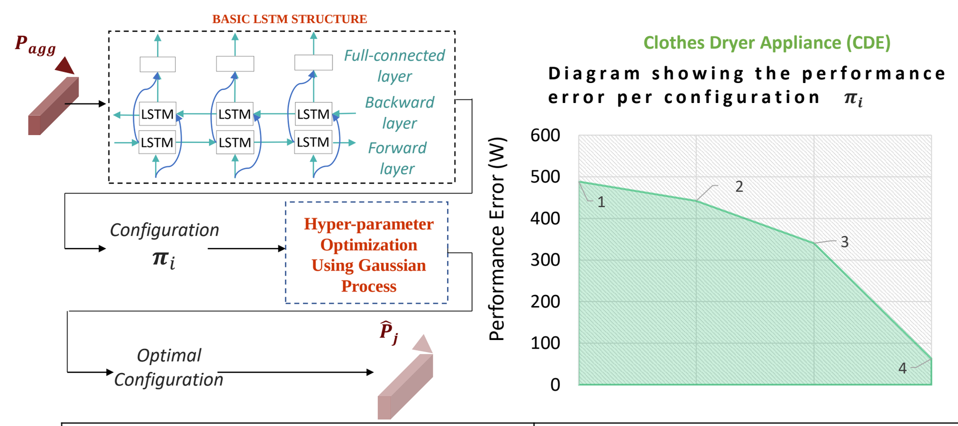


Fig.2 Shows signature identification examples for four selected appliances (CDE, DWE, HPE, WOE). Here, we illustrate the performance of our proposed method Bayes-BiLSTM compared to that of CNN (dotted line) as well as ground truth (grey filled).

Fig.2 shows signature identification examples for the four selected appliances. In this figure, we illustrate the performance of our proposed method compared to that of CNN as well as ground truth. As observed, our approach yields better performance in estimating not only ON/OFF states, but also more complicated energy patterns. This is important since it offers insight on whether a device is active as well as how it contributes to the total energy consumption.



Configurations π_i extracted from Bayesian Optimization	1	2	3	4
Number of BiLSTM layers	1	1	2	1
Hidden units in BiLSTM layers	74	77	42/34	60
Performance Error (RMSE (W))	488.67	442.04	340.55	62.19

Fig.3 Bayesian optimization results for CDE.

For each appliance, we have built a **Bayes-BiLSTM**. Bayesian optimization is used for optimally estimating the structure of each Bayes-BiLSTM, as illustrated in the flowchart showing the proposed methodology's adopted procedure (Fig.3). This figure also shows the CDE's validation performance and the model's hyperparameters respectively, for four successive iterations. The final iteration performs best, as expected.

Comparisons

Next, we perform comparisons using the Estimated Energy Fraction Index (EEFI) and Actual Energy Fraction Index (AEFI) indicators defined as:

$$EEFI(j) = \sqrt{\frac{\sum_{j=1}^{M} \hat{p}_j(n)}{\sum_n \sum_{j=1}^{M} \hat{p}_j(n)}} \qquad AEFI(j) = \sqrt{\frac{\sum_{j=1}^{M} p_j(n)}{\sum_n \sum_{j=1}^{M} \hat{p}_j(n)}}$$

$$DEFI(j) = |EEFI(j) - AEFI(j)|$$

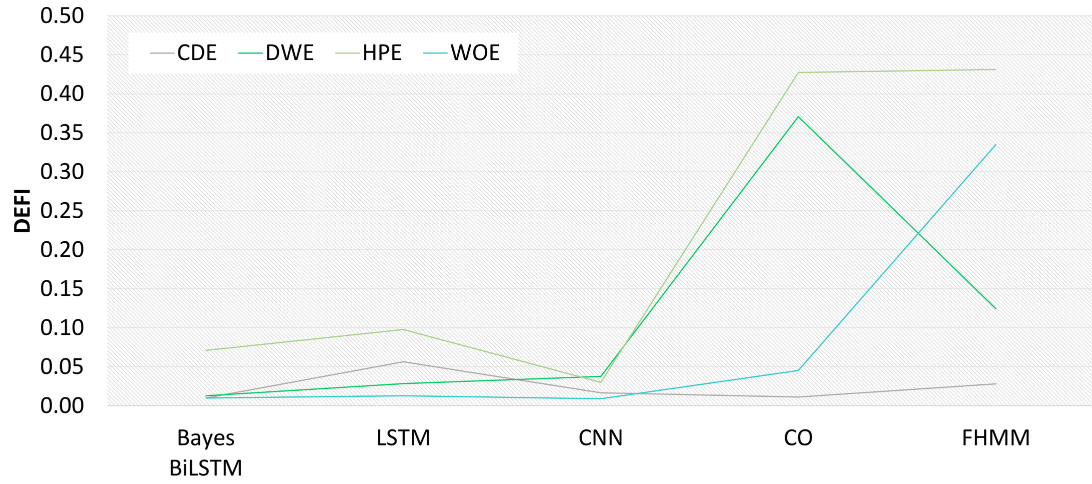


Fig.4 Comparisons using DEFI indicator.

Fig.4 Depicts the difference *DEFI* for the selected appliances over all compared methods, verifying that Bayes-BiLSTM yields the minimum value.

Conclusions

We propose a Bayesian-optimized Bidirectional LSTM regression model for NILM. The Bayes-BiLSTM model introduces: (i) a modular approach in NILM, which addresses dimensionality issues arising in cases of large number of appliances; (ii) a non-causal modeling framework taking into account the inherent structure, which characterizes the operation of multi-state appliances; (iii) a Bayesian optimization process ensuring the creation of a best fitting configuration for each appliance. Our proposed method is compared to the current state-of-the-art methods.

Acknowledgements

This research has received funding from the EU's H2020 research and innovation programme under grant agreement No768774.

References

[1] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Non-intrusive appliance load monitoring using low-resolution smart meter data," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2014, pp. 535–540.

[2] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[3] N. Batra et al., "NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring," *Proceedings of the 5th international conference on Future energy systems - e-Energy 14*, pp. 265–276, 2014.

[4] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. V. Baji, "AMPds: A public dataset for load disaggregation and eco-feedback research," in *2013 IEEE Electrical Power Energy Conference*, 2013, pp. 1–6.