

Short-Term Load Forecasting for Smart Home Appliances with Sequence to Sequence Learning

Mina Razghandi[†], Hao Zhou[‡], Melike Erol-Kantarci[‡], and Damla Turgut[†]

[†]Department of Computer Science, University of Central Florida

[‡]School of Electrical Engineering and Computer Science, University of Ottawa
mrazghandi@knights.ucf.edu, turgut@cs.ucf.edu, {hzhou098, melike.erolkantarci}@uottawa.ca

Abstract—Appliance-level load forecasting plays a critical role in residential energy management, besides having significant importance for ancillary services performed by the utilities. In this paper, we propose to use an LSTM-based sequence-to-sequence (seq2seq) learning model that can capture the load profiles of appliances. We use a real dataset collected from four residential buildings and compare our proposed scheme with three other techniques, namely VARMA, Dilated One Dimensional Convolutional Neural Network, and an LSTM model. The results show that the proposed LSTM-based seq2seq model outperforms other techniques in terms of prediction error in most cases.

Index Terms—load forecasting, deep learning

I. INTRODUCTION

Appliance-level load forecasting for residential consumers is of significant importance for the modern power system, particularly considering the dynamic loads such as PV power and EVs that will stress the distribution system and consequently call for fine-grained control of micro-loads by home energy management systems [1]. Therefore, mid-term and short-term load forecasting is essential for distribution system operation and control, while the long-term prediction plays a key role in infrastructure planning [2]. Short-term forecasting of household loads is closely related to consumer habits. An accurate and robust household load forecasting benefit the utility company and individual consumers. The utility company could better manage the electricity distribution and ancillary services while offering dynamic pricing to reduce peak demand. Meanwhile, individual consumers can better schedule the operation of their appliances to save from their electricity bills.

Electrical load forecasting is generally affected by many factors, e.g., such as weather conditions, social activity, consumers' habits, and so on. However, compared with load forecasting for a community or a commercial building, it is more challenging to forecast the short-term household load [3]–[7] due to the stochastic habits of consumers and the varying status and types of household appliances.

Several classification methods have been proposed in the literature for grouping similar load curves and eliminating the uncertainty [3]. The extraction of the consumption patterns of the electricity load can be accomplished by preprocessing the information using Fourier transforms or wavelet analysis [8].

The adoption of smart appliances by the consumers lead to changes to the traditional residential electrical load profiles [4].

To this end, modern AI techniques, e.g., Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Deep Belief Network (DBN), have been employed for household load forecasting [5]. Neural networks are capable of extracting the nonlinear and non-stationary nature of energy consumption. For instance, CNN has been used for short-term load forecasting in [6], [9], and the results show that CNNs outperform the Support Vector Machines (SVM), which is another machine learning algorithm. Furthermore, Long Short Term Memory (LSTM) network has been applied in [2], [7], [8], [10] for load forecasting, which presents a lower prediction error compared with traditional methods.

In this paper, we propose a novel LSTM-based Sequence to Sequence learning load forecasting method. While most previous studies mainly predict the overall household energy consumption, our work focuses on appliance-level short-term load forecasting, which has a higher level of uncertainty. Sequence to Sequence Learning (Seq2Seq) technique was initially proposed in [11], in which a multilayered LSTM network maps a sequence of input to a fixed-size vector and then, another multilayered LSTM network maps the fixed-sized vector to the sequence of target words. We use an LSTM network to map the sequence of past 24-hour energy consumption values to a fixed-size vector, then detect the appliance type, regenerate the input sequence in reverse form from the fixed-sized vector using another LSTM network, and produce a sequence of energy consumptions for the next hour from the fixed-length using another LSTM network. We show that Seq2Seq learning effectively deals with variable input and output sequence length and has more effective feature extraction abilities. We can see that our proposed model can distinguish the appliance type by its trend and learns appliances' typical usage duration. Our main contribution is the more accurate prediction of appliance-level consumption using a novel LSTM-based Seq2Seq learning algorithm.

The remainder of the paper is organized as follows. In Section II, we review previous work. We present our proposed LSTM-based Sequence to Sequence learning load forecasting method in Section III. The evaluation study provides detailed results in Section IV. Finally, Section V concludes the paper.

II. RELATED WORK

There have been a number of studies in the electric load forecasting area, and the typical prediction methods used include i) traditional machine learning (e.g., Linear Regression, SVM, AutoRegressive Integrated Moving Average), ii) deep learning algorithms (e.g., RNNs, CNNs, and LSTM networks), iii) probabilistic forecasting (e.g., Quantile Regression, Density Regression), and iv) hybrid methods (e.g., combining CNNs with Gated Recurrent Units).

Most recently, modern AI techniques, especially machine learning and deep learning, became the mainstream techniques for load forecasting [12]. A sparse coding-based learning approach for household electricity demand forecasting has been proposed by Yu et al. [13], and 10% higher accuracy has been observed compared to the existing studies. Based on multiple kernel learning, Wu et al. [14] proposed an improved gradient boosting framework, and the framework is extended to transfer learning context for both homogeneous and heterogeneous settings. Feng et al. [15] developed a two-step Q-learning, leveraging both deterministic and probabilistic load forecasting.

As stated earlier, recent years have witnessed the great potential of deep learning in many areas, which provides new opportunities for load forecasting as well. Bayesian deep learning has been applied in [16] to implement probabilistic load forecasting, and a clustering-based pooling method has been designed to prevent overfitting to improve the predictive performance. The overfitting problem has also been considered by Shi et al. [3], where a novel pooling deep RNN is proposed to learn the uncertainty of load forecasting directly. Deep learning has also been used in non-intrusive load monitoring where no sub-metered information is needed to estimate the demand of individual appliances [4], [17]. LSTM network has been applied by Kong et al. [2] for short-term residential load forecasting and has been shown to outperform the conventional neural network-based approaches.

There are also a few hybrid methods proposed. For instance, LSTM has been combined with stationary wavelet transform technique for individual household load forecasting in [8]. CNNs and Gated Recurrent Units have been unified in [9], resulted in lower computational complexity and higher prediction accuracy.

The inputs and targets should be encoded with vectors of fixed dimensionality to apply the deep neural network, which is a significant limitation. Different than prior works, in this paper, the sequential problem is solved by sequence to sequence learning. The idea of sequence to sequence learning is to use a multilayered LSTM to map the input sequence to a vector of a fixed dimensionality, and then another deep LSTM to decode the target sequence.

III. LOAD FORECASTING WITH SEQUENCE TO SEQUENCE LEARNING

In this paper, we use a Sequence to Sequence (Seq2Seq) encoder/decoder model to predict the energy consumption of smart home appliances, which represents the behavioral

pattern of smart home residents. The Seq2Seq encoder/decoder model derives occupants' habits from historical data and predicts hour-ahead appliance-level energy consumption in a household.

Sustkever et al. [11] first introduced the sequence to sequence learning technique in 2014. The idea was to encode a sequence of inputs to a fixed-length vector with a multilayered LSTM and then decode the fixed-length vector to a sequence of outputs with another multilayered , noting that input and output sequences length can vary. They used Seq2Seq learning to translate sentences from English to French.

We employed Seq2Seq learning to predict the energy consumption of smart home appliances for the next hour based on the previous 24 hours of historical data. As shown in Fig.1, our network architecture has three main LSTM network modules.

Considering (x_1, x_2, \dots, x_n) as the input sequence for the encoder module, the LSTM **encoder** network maps the input sequence to a fixed-length vector Z , which is mainly based on the last hidden state of LSTM. The initial hidden state of the LSTM network in the **decoder** module is set to the fixed-length vector Z and its goal is to compute and maximize the conditional probability of sequence $(x'_n, x'_{n-1}, \dots, x'_1)$ of (x_1, x_2, \dots, x_n) :

$$p(x'_n, \dots, x'_1 | x_1, \dots, x_n) = \prod_{t=1}^n p(x'_{n-t} | x'_n, \dots, x'_{n-t+1}, Z) \quad (1)$$

In the Equation 1, $(x'_n, x'_{n-1}, \dots, x'_1)$ is the reversed form of input sequence (x_1, x_2, \dots, x_n) enforcing encoder network to obtain long term dependencies of input sequence with putting more weight on most recent ones by minimizing a weighted MSE loss function (Equation 2). Besides this, by applying a Softmax function to the fixed-length vector Z , the decoder module learns the appliance working pattern and predicts the type of appliance that was not given in the input features.

$$WeightedMSE = \frac{1}{n} \sum_{i=1}^n W * (x'_i - x_i)^2 \quad (2)$$

Finally, the LSTM network in the **generator** module, estimates the conditional probability $p(y_1, \dots, y_m | x_1, \dots, x_n)$ according to the Equation 3:

$$p(y_1, \dots, y_m | x_1, \dots, x_n) = \prod_{t=1}^m p(y_t | y_1, \dots, y_{t-1}, Z) \quad (3)$$

In Equation 3, (y_1, \dots, y_m) is the prediction results for next m data points obtained from the fixed-length vector Z by minimizing a MSE loss function:

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (4)$$

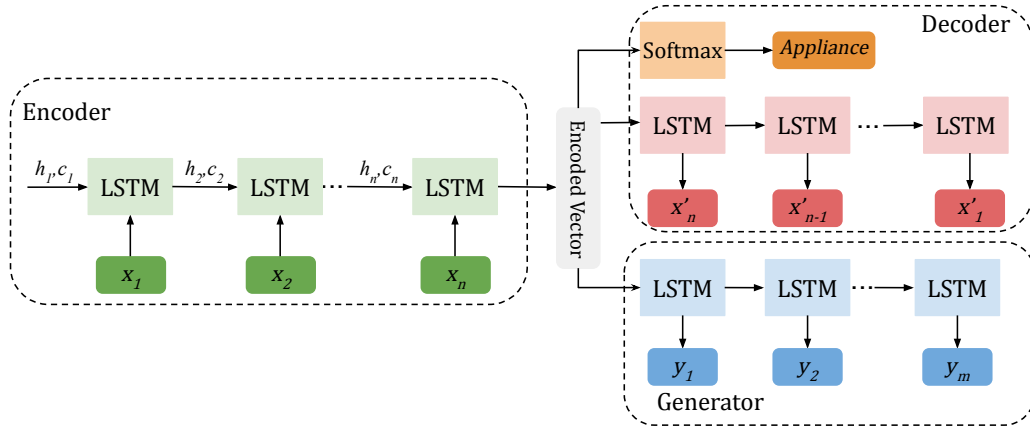


Fig. 1: Seq2Seq encoder decoder model.

TABLE I: Datasets used.

Building No.	Days	Monitored Appliances
Building 0	236	dishwasher, lamp, fridge, radio, washing machine
Building 1	226	dishwasher, lamp, fridge, radio, washing machine
Building 2	231	dishwasher, TV, network attached storage, laptop, washing machine
Building 3	151	dishwasher, TV, fridge, computer, washing machine

IV. EVALUATION STUDY

A. Experiment Setup

In the evaluation study, we used a publicly available dataset, GreenD [18], collected from eight different households, including apartments and detached houses, from December 2013 to October 2014 in Austria and Italy. After data-processing, only four buildings had sufficient data for training a deep learning model (more than 150 days) with $\frac{1}{60}$ Hz frequency. This dataset contains appliance-level energy consumption for an average of nine appliances per household at 1Hz frequency. However, not all of the monitored appliances display a repeated pattern of the occupant’s lifestyle. Therefore, Table I presents the five selected home appliances that were in common among all four buildings and had a repetitive trend. With 70%, 20%, and 10% ratio, we split data into three training, validation, and test sets.

We used three load forecasting methods from the literature as the baseline for evaluating the performance of the proposed model:

- **Vector Auto Regression Moving Average (VARMA):** This model is a combination of Vector Auto-Regression (VAR) and Vector Moving Average (VMA) models. VAR is a generalization of the Auto-Regressive model and used when there are multiple parallel variables to forecast. VAR models the variables as a linear function of their past variables, and the order of the VAR model determines the number of earlier time steps the model utilizes for predictions. VMA predicts the next steps based on a linear

function of residual errors, that is, the difference between predicted and observed values to predict.

- **Dilated One Dimensional Convolutional (CONV1D):** This model is based on the WaveNet network, proposed in [19] that has dilated causal convolution layers to transform text to speech. The 1-dimensional convolution slides a filter on an input series usually by one stride, but then in the causal convolution, the ordering of the input data will remain the same, and the model will not be learning based on future data. In a dilated convolution, the sliding filter skips input data with certain steps. Multiple stacked dilated convolutional layers allow larger input sequences but keep the network complexity efficient.
- **Long Short-Term Memory Neural Network:** LSTM network is a type of Recurrent Neural Network specifically designed to learn long-term dependencies in input data by relying on feedback from previous stage outputs. We used a 2-layer LSTM network in this paper.
- **Seq2Seq Learning:** As described in Section III.

B. Performance Indices

We have used the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics to evaluate regression prediction accuracy. However, since both RMSE and MAE are scale dependant, we have also considered Normalized Root Mean Square Error (NRMSE) in our experiments. RMSE, MAE and NRMSE are shown in Equations 5–7 where n is the total number of samples, y_j is the target value and \hat{y}_j is the predicted value, $\max(y_j)$ and $\min(y_j)$ refer to the maximum and minimum electrical usage recorded for the appliance j in the data, respectively.

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (y_j - \hat{y}_j)^2} \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_j - \hat{y}_j| \quad (6)$$

$$NRMSE = \frac{RMSE}{\max(y_j) - \min(y_j)} \quad (7)$$

C. Data Preprocessing

Linear interpolation is used to complete missing data. Moreover, instead of using minute-by-minute energy consumption of appliances, we smooth out the energy consumption of appliances for $T=10$ minutes without losing generality. In our dataset, appliance energy consumption varies between 10W to 2000W; therefore, the values are normalized before used in deep learning models.

Trends in energy consumption for electrical appliances can expose the living habits of household inhabitants. Some appliances would be used daily, such as TV or refrigerators, while others in a weekly fashion, such as vacuum cleaner, and some of them display no specific usage patterns. It is worth noting that appliances with repetitive usability patterns are more suitable candidates to make predictions based on them. Accordingly, the historical data of power consumption and the day of the week are selected as the network input features. The row index decides the minute and hour specification when the dataset is sorted in terms of time, so they are no longer needed as an additional feature input. Instead, for this model, the input sequence length of 144 points is chosen with T -minute intervals covering 24 hours.

D. Evaluation Results

Table II summarises RMSE, NRMSE, and MAE findings of four load forecasting algorithms for Building 0-3 with the lowest error in bold. Figure 2a–2f plot and compare the forecasting performance of all four prediction algorithms in a day snapshot against the ground truth for dishwasher, lamp, fridge, radio, laptop and TV.

The results demonstrate that the LSTM-based Seq2Seq and LSTM model, respectively, outperform the other forecasting models and have noticeably lower prediction error. LSTM model architecture is quite similar to the Seq2Seq model architecture (encoder, decoder, and generator). Consequently, the LSTM model has the closest results to the Seq2Seq model; however, the Seq2Seq model does a better job of learning the appliance's working pattern and detecting the appliance type. Unlike the LSTM model, the Seq2Seq model has no information about appliance type as an input.

As shown in Figure 2c, all four prediction models were effective in learning patterns for the appliance with the most seasonal energy usage pattern (fridge), which means that these models are more promising when there is a recurring pattern of behaviors of smart home residents. However, the Seq2Seq model performs better in load forecasting for appliances with a higher level of uncertainty in usage patterns.

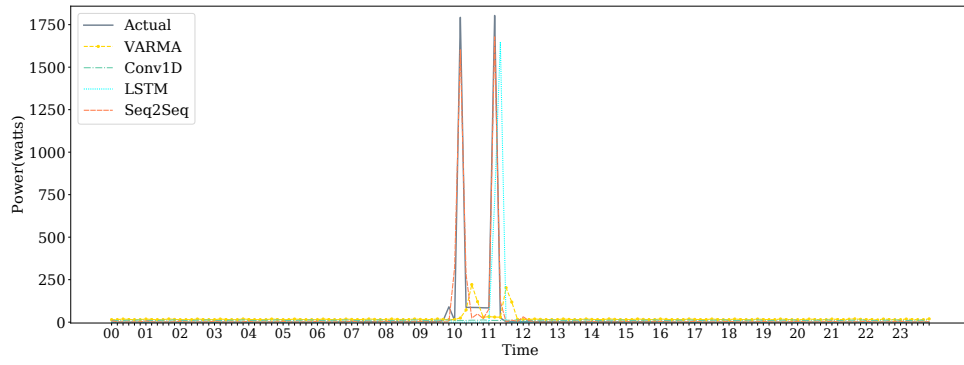
V. CONCLUSION

In this paper, we proposed an appliance-level load forecasting model for residential homes. Our proposed LSTM-based Sequence to Sequence learning model forecasts energy consumption for smart home appliances for the hour ahead with 10-minute resolution using historical data of the past day. We evaluated the performance of our model compared with the VARMA model, Dilated 1D Convolution, and LSTM

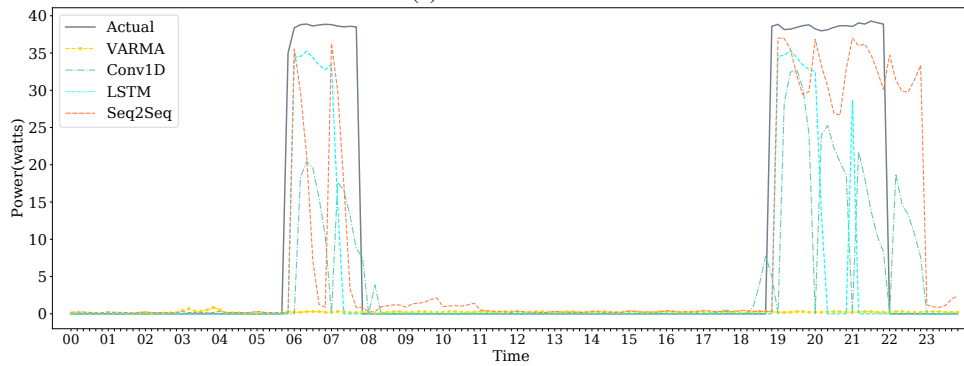
model on a publicly available dataset that contains historical data for multiple buildings. The consistent outperforming results demonstrate that the LSTM-based Sequence to Sequence learning model does not depend on the data and can be used for any household. As future work, we intend to combine optimization algorithms with prediction results to provide energy-efficient scheduling for a smart home.

REFERENCES

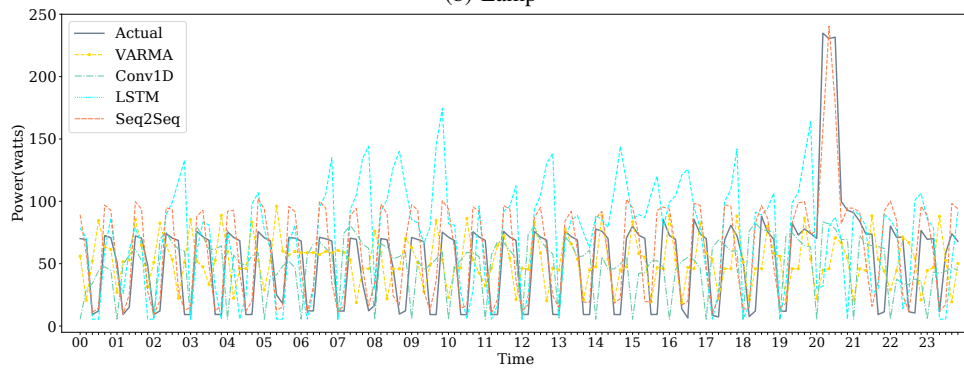
- [1] I. Zengin, J. S. Vardakas, J. Abadal, C. Echave, M. M. Güell, and C. Verikoukis, "Optimal power equipment sizing and management for cooperative buildings in microgrids," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 158–172, 2019.
- [2] W. Kong, Z. Dong, Y. Jia, D. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. on Smart Grid*, vol. 10, no. 1, pp. 841–85, Jan. 2019.
- [3] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—a novel pooling deep RNN," *IEEE Trans. on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.
- [4] W. Kong, Z. Dong, B. Wang, J. Zhao, and J. Huang, "A practical solution for non-intrusive type II load monitoring based on deep learning and post-processing," *IEEE Trans. on Smart Grid*, vol. 11, no. 1, p. 148–160, Jan. 2020.
- [5] K. Amarasinghe, D. L. Marino, and M. Manic, "Deep neural networks for energy load forecasting," in *Proc. of IEEE Int. Symp. on Industrial Electronics*, Jun. 2017, pp. 1–6.
- [6] A. Almalaq and G. Edwards, "A review of deep learning methods applied on load forecasting," in *Proc. of IEEE Int. Conf. on Machine Learning and Applications*, Dec. 2017, pp. 1–6.
- [7] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network," in *Proc. of Annual Conf. on Information Sciences and Systems*, Mar. 2017, pp. 1–6.
- [8] K. Yan, W. Li, Z. Ji., M. Qi, and Y. Du., "A hybrid LSTM neural network for energy consumption forecasting of individual households," *IEEE Access*, vol. 7, pp. 157 633–157 642, 2019.
- [9] M. Sajjad, A. Khan, A. Ullah, T. Hussain, Q. Ullah, M. Lee, and S. Baik, "A novel CNN-GRU-based hybrid approach for short-term residential load forecasting," *IEEE Access*, vol. 8, pp. 143 759–143 768, 2020.
- [10] M. Razghandi and D. Turgut, "Residential appliance-level load forecasting with deep learning," in *Proc. of IEEE GLOBECOM*, Dec. 2020, pp. 1–6.
- [11] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. of Int. Conf. on Neural Information Processing Systems*, 2014, pp. 3104–3112.
- [12] K. Amasyali and N. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renewable Sustainable Energy Reviews*, vol. 81, pp. 1192–1205, Jan. 2019.
- [13] C. Yu, P. Mirowski, and T. Ho, "A sparse coding approach to household electricity demand forecasting in smart grids," *IEEE Trans. on Smart Grid*, vol. 8, no. 2, pp. 738–748, Mar. 2017.
- [14] D. Wu, B. Wang, D. Precup, and B. Boulet, "Multiple kernel learning based transfer regression for electric load forecasting," *IEEE Trans. on Smart Grid*, vol. 11, no. 2, pp. 1183–1192, Aug. 2019.
- [15] C. Feng, M. Sun, and J. Zhang, "Reinforced deterministic and probabilistic load forecasting via Q-learning dynamic model selection," *IEEE Trans. on Smart Grid*, vol. 11, no. 2, pp. 1377–1386, Aug. 2019.
- [16] Y. Yang, W. Li, T. Gulliver, and S. Li, "Bayesian deep learning-based probabilistic load forecasting in smart grids," *IEEE Trans. on Industrial Informatics*, vol. 16, no. 7, pp. 4703–4713, Jul. 2020.
- [17] N. Linh and P. Arbolea, "Deep learning application to non-intrusive load monitoring," in *Proc. of IEEE Milan PowerTech*, Jun. 2019, pp. 1–6.
- [18] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, and A. M. Tonello, "GREEN: an energy consumption dataset of households in Italy and Austria," in *Proc. of SmartGridComm*, 2014, pp. 511–516.
- [19] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," 2016.



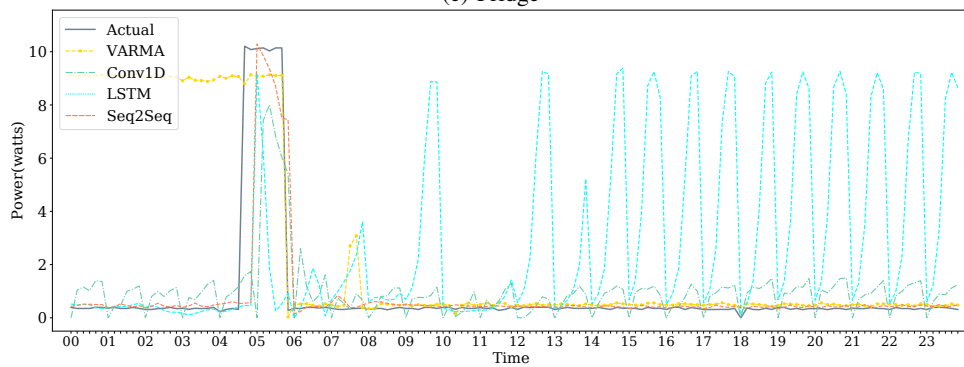
(a) Dish Washer



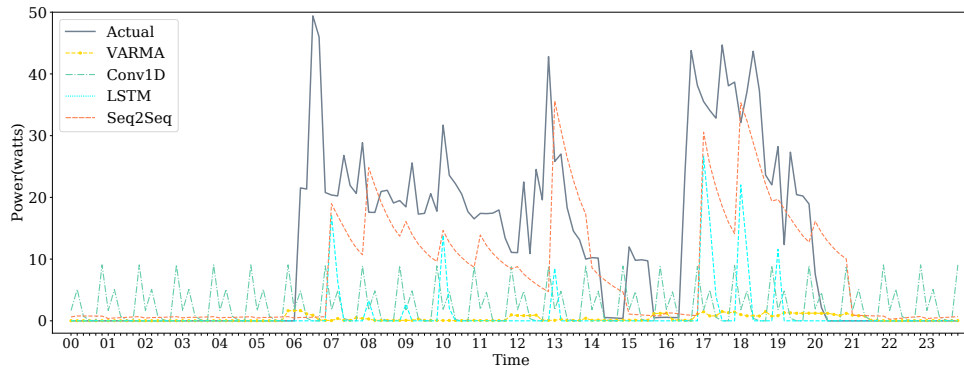
(b) Lamp



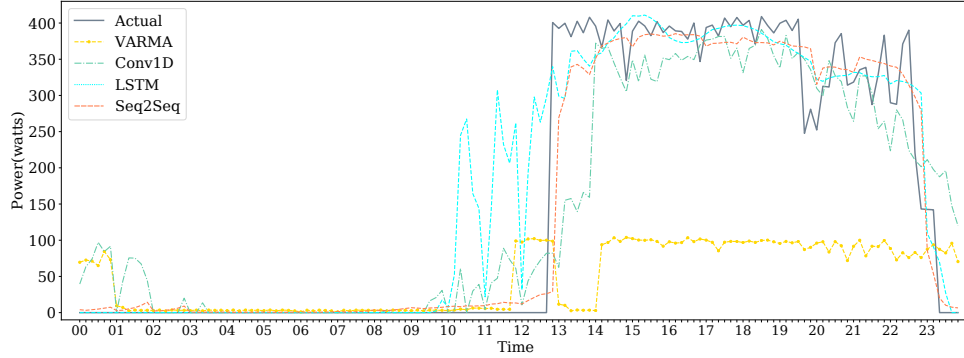
(c) Fridge



(d) Radio



(e) Laptop



(f) TV

Fig. 2: Energy consumption prediction for 24 hours with a 10-minute resolution.

TABLE II: Evaluation Results.

Building 0															
Model	Dish Washer			Lamp			Fridge			Radio			Washing Machine		
	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE
VARMA	229.237	0.111	53.722	18.452	0.450	9.456	44.993	0.179	34.197	4.191	0.408	2.056	236.026	0.112	64.035
Conv1D	198.400	0.096	39.421	9.837	0.240	3.842	46.133	0.183	34.073	2.395	0.233	1.156	214.037	0.102	51.723
LSTM	493.948	0.239	168.236	11.632	0.284	3.942	46.601	0.185	33.218	3.513	0.342	1.805	246.735	0.117	46.816
Seq2Seq	189.799	0.092	37.394	9.185	0.224	3.623	43.748	0.174	30.587	2.314	0.225	0.954	215.298	0.102	48.373
Building 1															
Model	Dish Washer			Lamp			Fridge			Radio			Washing Machine		
	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE
VARMA	166.513	0.091	36.364	17.736	0.168	6.140	39.480	0.330	31.445	4.615	0.246	1.766	140.496	0.069	23.545
Conv1D	164.924	0.090	27.609	14.757	0.140	5.298	25.534	0.213	15.541	2.860	0.152	0.867	129.324	0.063	11.285
LSTM	154.280	0.084	18.174	11.855	0.112	2.875	36.070	0.301	20.999	2.211	0.118	0.351	120.479	0.059	33.571
Seq2Seq	119.308	0.065	13.782	12.803	0.121	3.420	23.346	0.195	10.920	2.314	0.123	0.420	119.215	0.058	11.960
Building 2															
Model	Dish Washer			TV			Network attached storage			Laptop			Washing Machine		
	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE
VARMA	294.618	0.135	86.852	203.580	0.478	141.565	79.463	0.550	69.964	16.122	0.030	3.785	138.297	0.112	23.292
Conv1D	294.417	0.134	73.202	116.982	0.275	68.136	10.406	0.072	3.808	16.236	0.030	6.329	137.647	0.111	43.017
LSTM	299.399	0.137	63.336	124.385	0.292	61.287	7.906	0.055	2.418	15.800	0.029	3.364	177.153	0.143	98.461
Seq2Seq	256.000	0.117	69.832	113.047	0.266	54.230	9.307	0.064	2.450	14.537	0.027	3.061	142.829	0.115	34.219
Building 3															
Model	Dish Washer			TV			Fridge			Computer			Washing Machine		
	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE	RMSE	NRMSE	MAE
VARMA	157.059	0.088	29.454	34.862	0.363	14.512	70.842	0.558	49.705	38.798	0.433	25.994	189.403	0.095	39.776
Conv1D	152.446	0.085	16.818	22.188	0.231	8.625	41.075	0.323	24.815	15.416	0.172	8.242	165.717	0.083	20.852
LSTM	135.093	0.076	12.546	18.148	0.189	4.666	46.150	0.363	25.728	14.639	0.164	5.947	187.217	0.094	40.206
Seq2Seq	131.433	0.073	18.913	21.599	0.225	7.167	32.821	0.258	17.385	15.813	0.177	7.124	164.082	0.082	31.318