An LSTM-Based Online Prediction Method for Building Electric Load During COVID-19

Hao Tu[1], Srdjan Lukic[2], Abhishek Dubey[3] and Gabor Karsai[4]

[1,2] *FREEDM Systems Center, North Carolina State University, Raleigh, NC, 27606, USA*
htu@ncsu.edu
smlukic@ncsu.edu

[3,4] *Institute for Software Integrated Systems, Vanderbilt University, Nashville, TN, 37212, USA*
abhishek.dubey@vanderbilt.edu
gabor.karsai@vanderbilt.edu

ABSTRACT

Accurate prediction of electric load is critical to optimally controlling and operating buildings. It provides the opportunity to reduce building energy consumption and to implement advanced functionalities such as demand response in the context of the Smart Grid. However, buildings are non-stationary and it is important to consider the underlying concept changes that will affect the load pattern. In this paper we present an online learning method for predicting building electric load during concept changes such as COVID-19. The proposed methods is based on online Long Short-Term Memory (LSTM) recurrent neural network. To speed up the learning process during concept changes and improve prediction accuracy, an ensemble of multiple models with different learning rates is used. The learning rates are updated in real-time to best adapt to the new concept while maintaining the learned information for the prediction.

1. INTRODUCTION

Buildings consume a large portion of electricity today. Accurate predicting of building electric load can help the building management system to schedule the building operation more efficiently. With the advance of the Smart Grid concept (Tu et al., 2020), buildings can interact with other grid components and become an active player that can help support grid operation such as by providing frequency regulation service (Goddard, Klose, & Backhaus, 2014; Beil, Hiskens, & Backhaus, 2016). If a building is equipped with photovoltaic generation systems and battery energy storage systems, it can inject power back into the grid when the energy price is high and recharge the battery storage when the energy cost is low (Y. Wang, Wang, Chu, Pota, & Gadh, 2016). Some other researchers also look at building in microgrids (Guan, Xu, & Jia, 2010) and study energy management strategies of maintaining self-sufficient operation when the main grid is disconnected from the building microgrid. An accurate predicting method of building electric load is critical to realizing all the advanced functionalities.

Several building modeling methods have been proposed in literature to predict buildings' energy use. They can be briefly categorized in white-box, grey-box, and black-box approaches (Li & Wen, 2014). The white-box approaches run detailed physical models of the buildings and predicts the energy use based on the underlying law of physics. However, this approach requires the properties of the buildings such as geometric shape and wall materials as the input and they are generally hard to obtain. The grey-box approach uses simplified models to simulate the energy behavior buildings. For example, an resistor-capacitor network can be used to model the building's thermal behavior (S. Wang & Xu, 2006). For grey-box approach, the parameter identification for the simplified models is critical to the prediction accuracy and requires careful calibration. The black-box approach uses data-driven methods to make prediction. Instead of constructing physical models of buildings, historical data are used to find the energy consumption patterns and make prediction based on the recognized patterns. For example, an auto-regressive model (Yun, Luck, Mago, & Cho, 2012) with exogenous inputs such as temperature, solar irradiance and wind speed is proposed. Recently, machine learning based methods are widely studied to predict the building energy use. For example, a radial basis function neural network (Mai, Chung, Wu, & Huang, 2014) or Long Short-Term Memory (LSTM) Network (Marino, Amarasinghe, & Manic, 2016) can be used to predict the building energy demand.

One major challenge associated with the machine learning based methods is concept change. Concept change refers to the "changes in the conditional distribution of output given the input" (Gama, Žliobaitė, Bifet, Pechenizkiy, & Bouchachia, 2014). More specifically, the machine learning based methods predicts the future energy consumption using historical data. This assumes the future energy use will follow the same pattern learned from the historical data. However, this assumption may not be always valid. The building energy use pattern may varies in time due to different reasons. When the learned pattern does not reflect the current energy use pattern, the prediction made can be erroneous.

One way of dealing concept change is by employing online learning which retrains the machine learning model with new incoming data in every time step. Online learning has the capability of adapting to the new concept and its adaptation speed have direct impact on the prediction accuracy. One method called Dynamic Weighted Majority (DWM) was proposed to speed up adaptation and improve prediction accuracy (Kolter & Maloof, 2007). DWM maintains an ensemble of multiple models and the final prediction of the ensemble is the weighted average of all models. The weight of a particular model depends on its history of prediction accuracy. Another approach of adjusting the adaption speed is to use an adaptive learning rate in online training (Guo et al., 2016; Yang, Pan, & Tao, 2017). Both papers aims at reducing the adverse effect of outliers by tuning down the learning rate when an outlier is detected. An online learning method is proposed for anomaly detection and concept change using multi-step prediction (Saurav et al., 2018). The proposed method with local normalization can quickly adapt to certain types of changes such as mean value and frequency but fail to address concept changes in general.

This paper proposes an LSTM-based online prediction method. The proposed method maintains an ensemble of multiple models with different learning rates. The learning rates are adaptive depending on the prediction accuracy of the models in the ensemble. The proposed method is tested on a building electric load dataset during COVID-19, showing that it can quickly adapt to new concepts and make predictions with improved accuracy.

2. LONG SHORT TERM MEMORY FOR BUILDING ELECTRIC LOAD PREDICTION

2.1. Long Short-Term Memory and Single-Step Prediction

LSTM is a type of recurrent neural network that is able to learn long-term dependency from data and it is particularly suitable for applications such as time series prediction. More detailed information about LSTM can be found in (Hochreiter & Schmidhuber, 1997). Input to LSTM can be a historical

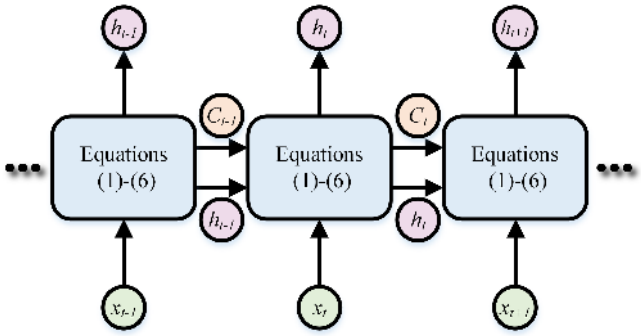


Figure 1. LSTM layer

sequence $\{x_k, x_{k-1}, ..., x_{k-l}\}$ with length l. The LSTM cell for a recurrent step can be described by equations (1)-(6).

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$
$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$
$$\tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \quad (4)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$
$$h_t = o_t * \tanh(C_t) \quad (6)$$

Where $W_i, U_i, b_i, W_f, U_f, b_f, W_o, U_o, b_o, W_C, U_C, b_C$ are the trainable parameters of the LSTM cell; σ is the sigmoid activation; and * denotes element-wise multiplication. The input for the tth ($t \leq k$) recurrent step includes external input x_{t-1} and the carry state C_{t-1} and hidden state h_{t-1} from the previous recurrent step as shown in Figure 1. The update of carry state C_t is governed by input gate i_t and forget gate f_t. The output (and also its hidden state) h_t of the recurrent step is calculated based on carry state C_t and output gate o_t. The hidden states $h_t (1 \leq t \leq k)$ can be used as the input to the next LSTM layer or to calculate the final output of the model.

For single-step time series prediction, the historical sequence $\{x_k, x_{k-1}, ..., x_{k-l}\}$ are used to make prediction for the next time step $\tilde{y}_{k+1}$. After the LSTM layers, a dense layer is added to make the target prediction based on the output of the last LSTM recurrent step. Thus, the final prediction is given as,

$$\tilde{y}_{k+1} = W_d h_k + b_d \quad (7)$$

Where W_d, b_d are the trainable parameters of the dense layer.

2.2. LSTM for Building Electric Load Prediction

For the application of building load prediction, while the target prediction $\tilde{y}_{k+1}$ can be easily identified as the electric load level for the next time step, several features can be identified as input to the LSTM model.

1. As a significant portion of the electric load is consumed

by the building's heating, ventilation, and air conditioning (HVAC) systems, the building electric load is closely related to the outside air temperature (OAT), T_{oat}. Thus, the current OAT is used as an input feature.

2. The building electric load varies with the time of the day. Generally, the building load peak is around 2 PM while the load during night is very low. This feature is a vector $H \in \{0,1\}^{24}$, presenting the hour index of the day using one hot encoding. For example, 2 AM will be encoded as a 24-dimensional vector whose third element is 1 and all the other elements are zero.
3. The building electric load varies with the day of the week. For example, the load during weekdays is generally higher than weekends. This feature is a vector $D \in \{0,1\}^{7}$, presenting the day index of the week using one hot encoding. For example, a Monday will be encoded as a 7-dimensional vector whose first element is 1 and all the other elements are zero.
4. The building electric load is significantly lower in holidays than that in work days. A binary feature G is used to denote whether it is a holiday or not, i.e. $G = 1$ if it is a holiday and $G = 0$ otherwise.
5. The above features capture some critical information that can affect the building electric load. However, other information like the occupancy of the building, solar irradiance and wind speed is difficult to obtain and thus difficult to be included in the features explicitly, although they can influence the building power consumption. To include their influence, the current electric load y_k is used as input feature to predict the electric load for the next time step $\tilde{y}_{k+1}$.

For each time step k, all the features for that time step are concatenated into one feature vector $x_k = \{y_k, T_{oat,k}, H_k, D_k, G_k\} \in \mathbb{R}^{34}$. The LSTM layer takes a sequence of the feature vectors as input. The length of the sequence is selected to be 24 in this paper assuming that the load for the next hour can be predicted based on the information from the last 24 hours. Thus, the input to the LSTM is sequence $\{x_k, x_{k-1}, ..., x_{k-23}\}$ and the output is the predicted electric load for the next time step $\tilde{y}_{k+1}$.

3. ONLINE LEARNING WITH ADAPTIVE LEARNING RATE

3.1. Online Learning

Online learning is able to update an existing model using only the new incoming data without retraining it on the entire data set. It provides a computationally efficient way of handling new data. More importantly, online learning has the capability of adapting the model to new incoming data and discovering the underlying new pattern if a concept change has happened.

Algorithm 1 Online LSTM with Adaptive Learning Rate

procedure PredictAndUpdate
  for $i = 1 \rightarrow 3$ **do**
    $\{\tilde{y}_{k+1}^{(i)}\} \leftarrow$ PredictLearner$_i(x_k, ...x_{k-l}, \mathbf{W}_k^{(i)})$
  $\tilde{y}_{k+1} = (\tilde{y}_{k+1}^{(1)} + \tilde{y}_{k+1}^{(2)} + \tilde{y}_{k+1}^{(3)})/3$

  while (y_{k+1} is not available) **do**
    Wait

  for $i = 1 \rightarrow 3$ **do**
    $err^{(i)} = |\tilde{y}_{k+1}^{(i)} - y_{k+1}|$
  IdxBestLearner $\leftarrow$ index for lowest error $err^{(i)}$

  if *IdxBestLearner* == 1 **then**
    $\{\alpha_1, \alpha_2, \alpha_3\} \leftarrow \{\alpha_1 - \delta\alpha, \alpha_2 - \delta\alpha, \alpha_3 - \delta\alpha\}$
    BestLearner $\leftarrow$ *SlowLearner*
  else if *IdxBestLearner* == 2 **then**
    $\{\alpha_1, \alpha_2, \alpha_3\} \leftarrow \{\alpha_1, \alpha_2, \alpha_3\}$
    BestLearner $\leftarrow$ *AverageLearner*
  else if *IdxBestLearner* == 3 **then**
    $\{\alpha_1, \alpha_2, \alpha_3\} \leftarrow \{\alpha_1 + \delta\alpha, \alpha_2 + \delta\alpha, \alpha_3 + \delta\alpha\}$
    BestLearner $\leftarrow$ *FastLearner*
  {*SlowLearner, AverageLearner, FastLearner*} $\leftarrow$ {*BestLearner, BestLearner, BestLearner*}

  for $i = 1 \rightarrow 3$ **do**
    TrainLearner$_i$ on $(x_k, ...x_{k-l}, y_{k+1})$ with α_i

The online learning process for LSTM is illustrated as follow. For each time step k, the model makes a prediction $\tilde{y}_{k+1}$ based on the input $\{x_k, x_{k-1}, ..., x_{k-23}\}$ and model parameters $\mathbf{W}_k = \{W_i, U_i, b_i, W_f, U_f, b_f, W_o, U_o, b_o, W_C, U_C, b_C, W_d, b_d\}_k$ at that time step. When the ground truth y_{k+1} comes at time step $k+1$, a loss function is used to evaluate the error between the prediction $\tilde{y}_{k+1}$ and actual value y_{k+1}. The parameters $\mathbf{W}_k$ is then updated based on gradient descent method,

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \alpha \nabla loss(\mathbf{W}_k, y_{k+1}, \tilde{y}_{k+1}) \tag{8}$$

Where $loss$ is the selected loss function and α is the learning rate. With incoming ground truth data y_{k+1}, the feature vector for time step $k+1$, $x_{k+1} = \{y_{k+1}, T_{oat,k+1}, H_{k+1}, D_{k+1}, G_{k+1}\}$, and the new input sequence, $\{x_{k+1}, x_k, ..., x_{k-22}\}$, can be constructed. The new input sequence is fed to the updated model with parameters $\mathbf{W}_{k+1}$ to make the prediction $\tilde{y}_{k+2}$ for the next time step.

3.2. Online Learning with Adaptive Learning Rate

Learning rate α is critical to online learning algorithms. On one hand, a small learning rate results in a low learning speed thus requires more time steps to adapt to the new concept. The predictions during the adaption can suffer from poor accuracy. On the other hand, if the learning rate is too large, the network weights are changed significantly during each time step. Since the measurement of each time step may contain noises, the noises propagate to the network weights through

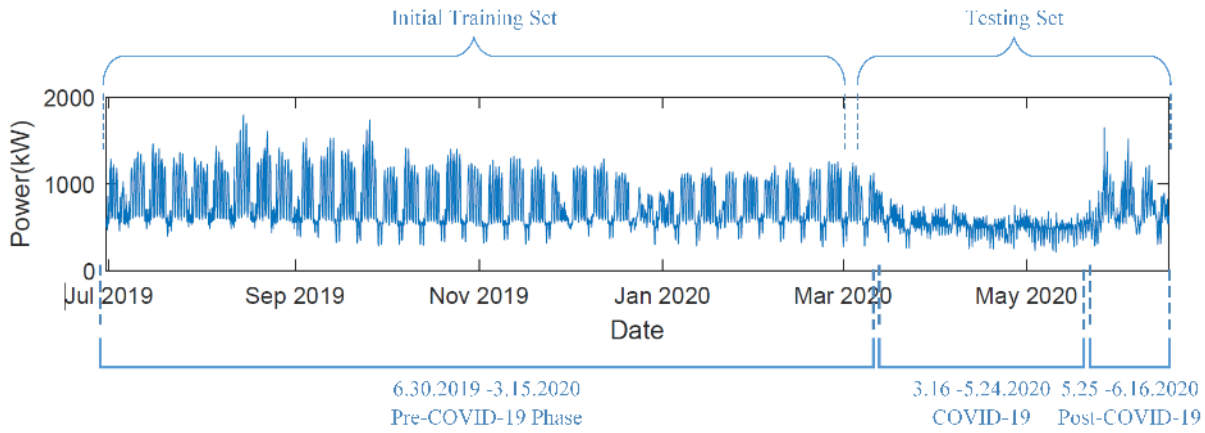


Figure 2. Hourly electrical load data of an office building from Jun 30 2019 to Jun 17 2020

Phase	Pre-COVID-19	COVID-19					Post-COVID-19		
Week	3.9-3.15	3.16-3.22	3.23-3.29	5.4-5.10	5.11-5.17	5.18-5.24	5.25-5.31	6.1 - 6.7	6.8-6.14
Off-LSTM	31.5	34.3	35.9	43.0	41.3	54.8	76.2	62.0	59.1
On-LSTM-FLR	35.0	35.5	31.3	33.4	34.9	51.3	69.1	59.7	53.0
Proposed	31.7	30.9	30.5	31.4	32.4	44.4	67.5	55.9	51.5

Table 1. MAE in kW for different methods

the learning rate and result in unstable predictions and large errors.

As it is difficult to determine the best learning rate beforehand, we propose to maintain three LSTM models (referred to as learners) with different learning rates $\alpha_1 < \alpha_2 < \alpha_3$ in a model ensemble. The learner with smallest learning rate α_1 is referred to as slow learner while the one with the largest learning rate α_3 as fast learner. The learner in the middle is referred to as average learner. At each time step, each learner makes a prediction and the final prediction is the average of the predictions made by all learners.

When the ground truth comes, the prediction made by individual learner is compared with the sample. If the fast learner demonstrates the lowest prediction error in all three learners, the learning rates of all learners are increased by a preset value $\delta\alpha$. If the slow learner has the lowest error, the learning rates of all learners are decreased by $\delta\alpha$. If the average learner has the lowest error, the learning rates are unchanged for this time step. Then, the learners in the ensemble are replaced by the learner that has the lowest prediction error. This is done by removing the all learners except the one with the lowest error and duplicated it twice. As a result, all the learners are identical (and they are the learner with the lowest prediction error for the last time step). Finally, the learners are trained on the new sample with the updated learning rates and make predictions for the next time step. Again, the final prediction is the average prediction of all learners.

The predicting and updating procedure is summarized in Algorithm 1.

4. RESULTS AND DISCUSSIONS

4.1. Data analysis

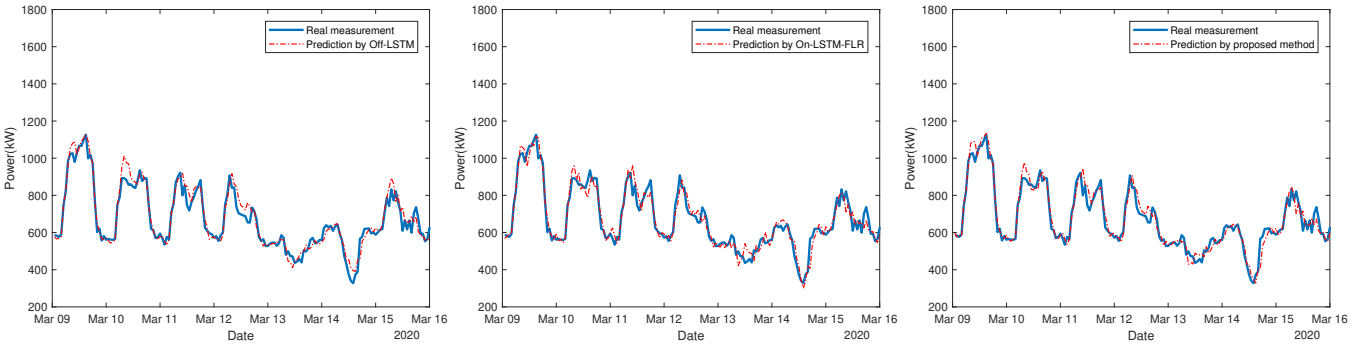
The electric load data of an office building in California, USA from 6.30.2019 to 6.17.2020 were collected. As shown in Figure 2, the data can be divided into three phases. From 6.30.2019 to 3.15.2020, it is the pre-COVID-19 phase when the building operation is normal and its electric load follows some regular pattern. From 3.16.2020 to 5.24.2020, it is the COVID-19 phase when the building is closed and only some essential departments were operating. The electric load drops significantly during the COVID-19 phase and its pattern differs from that of pre-COVID-19 phase. From 5.25.2020 to 6.17.2020, it is post-COVID-19 phase and the building is reopen. The building electric load increases during this phase. However, its pattern is not the same as the COVID-19 phase nor the pre-COVID-19 phase.

Based on above discussion, two concept changes are identified. The first one happens at 3.16.2020 when the building is closed. The second one happens at 5.25.2020 when the building is re-opened.

4.2. Baselines

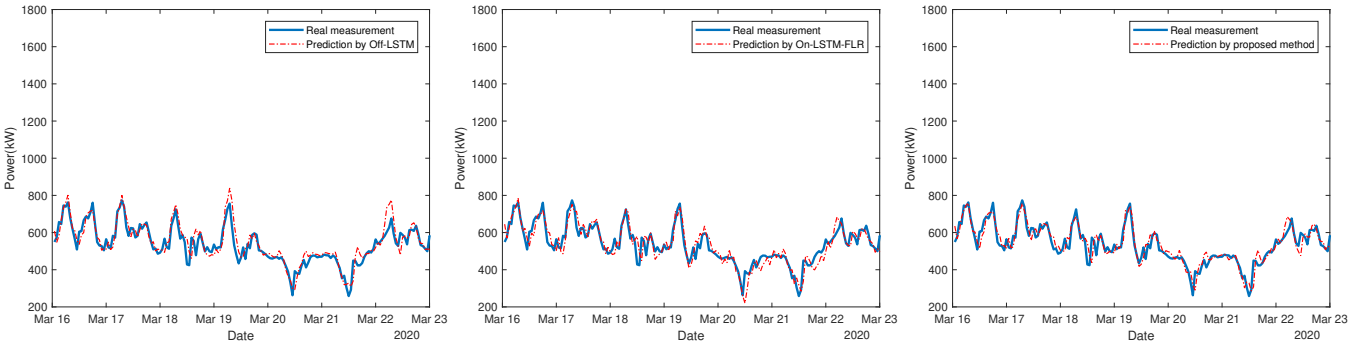
Two baselines are compared with the proposed method.

1. Off-line LSTM (Off-LSTM): this is LSTM model without online training. In this case, the data from 6.30.2019 to 3.5.2020 are used for training.
2. Online LSTM with fixed learning rate (On-LSTM-FLR): this methods takes the trained off-line LSTM model and



(a) Off-LSTM, MAE = 31.5 kW (b) On-LSTM-FLR, MAE = 35.0 kW (c) Proposed method, MAE = 31.7 kW

Figure 3. Results for pre-COVID-19 phase, week from Mar 9 2020 to Mar 15 2020



(a) Off-LSTM, MAE = 34.3 kW (b) On-LSTM-FLR, MAE = 35.5 kW (c) Proposed method, MAE = 30.9 kW

Figure 4. Results for COVID-19 phase, week from Mar 16 2020 to Mar 22 2020

update the model on new samples every time step with a fixed learning rate $\alpha = 0.02$.

Similar to On-LSTM-FLR, the proposed online LSTM with adaptive learning rate, takes the trained off-line LSTM model and applied the updating and training rules proposed in section 3. The proposed method and two baselines use the the same network structure. To be specific, two LSTM layers are stacked. The first LSTM layer has 32 neurons while the second layer has 16 neurons. Recurrent dropout rate is set to 50% for both LSTM layers. A dropout layer with 20% dropout rate is inserted between the two LSTM layers. The initial learning rates for the proposed method are set to $\alpha_1 = 0.01, \alpha_2 = 0.012, \alpha_3 = 0.014$. The learning rate increment is set to $\delta\alpha = 0.0002$.

The performance metric used to evaluate different models is mean absolute error (MAE),

$$\text{MAE} = \frac{1}{N} \sum_{k=N_0}^{N_0+N} |\tilde{y}_k - y_k| \tag{9}$$

Where N_0 is the starting time step and N is the number of predictions. As the proposed method in this paper aims at improving the prediction accuracy during concept change, N is selected to 168 (which corresponds to one-week long hourly data) to evaluate its performance change over time.

4.3. Test Results

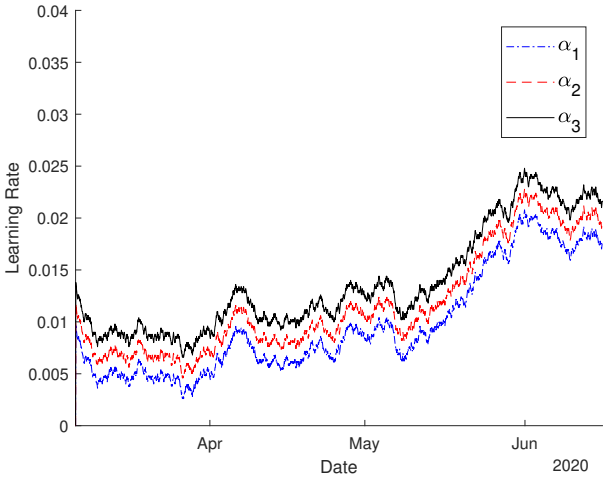
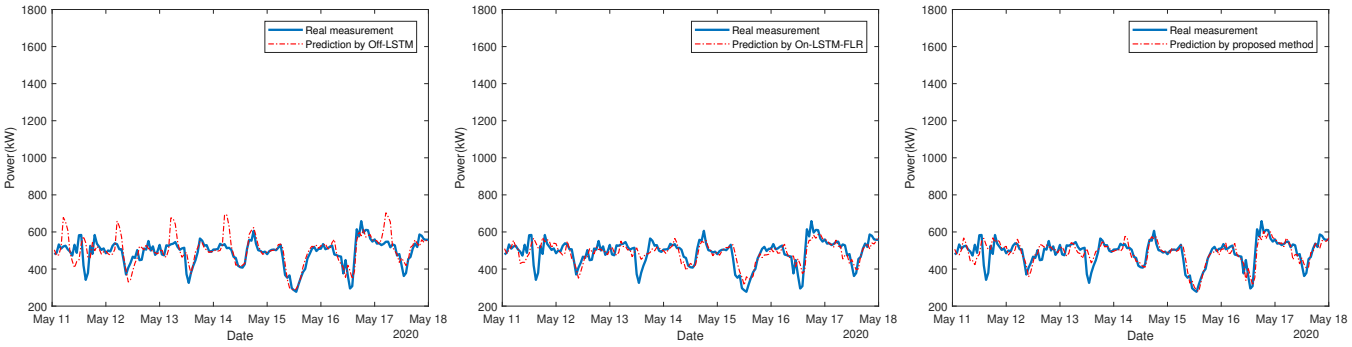


Figure 6. learning rate evolution of the proposed method from 3.6.2020 to 6.16.2020



(a) Off-LSTM, MAE = 41.3 kW (b) On-LSTM-FLR, MAE = 34.9 kW (c) Proposed method, MAE = 32.4 kW

Figure 5. Results for COVID-19 phase, week from May 11 2020 to May 17 2020

The performance of the models is tested on fourteen weeks across three phases as shown in Figure 2. The results are summarized in Table 1. It is worth noting the results during 3.30.2020-5.3.2020 are omitted because of space limitations. The results during the omitted period show similar trend with the proposed method having the lowest MAE and Off-LSTM having the highest MAE.

The learning rate evolution of the proposed method from 3.6.2020 to 6.16.2020 is shown in Figure 6.

4.3.1. Pre-COVID-19 Phase

Week-long data from 3.9.2020 to 3.15.2020 were used to test the performance of the three models during pre-COVID-19 Phase. As shown in Figure 3, Off-LSTM and the proposed method give similar performance on this test set while On-LSTM-FLR performing worse than them. Since the load during this week still follows the pattern of the building's normal operation, Off-LSTM can predict the load consumption used the learned off-line model with the lowest MAE. For online algorithms (On-LSTM-FLR and proposed method), the models are updated with new samples in each time step. The random noises in the new samples can disturb the learned weights and negatively influence the prediction. As the proposed method uses an adaptive learning rate, the learning rate is reduced during this phase to minimize the negative influence of the noise. This test shows that, first, the off-line LSTM model is properly trained to predict the building load when there is no concept change; second, the proposed method outperforms On-LSTM-FLR when there is no concept change.

4.3.2. COVID-19 Phase

During COVID-19, the building was closed thus its electric load was reduced significantly. The load pattern is also different from that of the pre-COVID-19 phase. The test results for two weeks during the COVID-19 phase are presented.

The results for the week from 3.16.2020 to 3.22.2020 is shown in Figure 4. This is the immediate week after the building was closed and corresponds to the first concept change. The proposed method shows the lowest MAE among the three methods, demonstrating its capability of adapting to the new concept and rejecting the influences of noises. Off-LSTM outperforms On-LSTM-FLR but the difference in MAE is smaller than that during pre-COVID-19. This is maybe because the new concept (new electric load pattern) during COVID-19 phase has not been stabilized in the first week. As a result, the pattern in this week presents some similarity to that of the pre-COVID-19 phase.

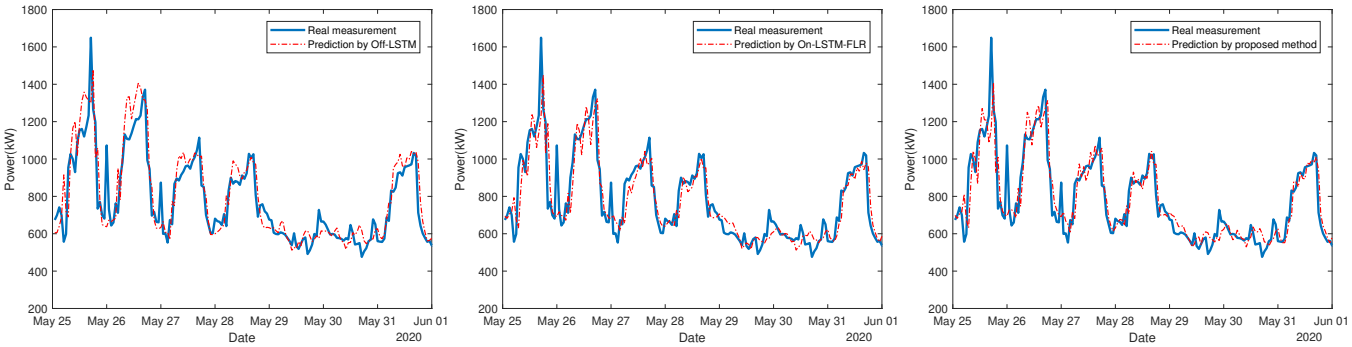
The results for the week from 5.11.2020 to 5.17.2020 is shown in Figure 5. This is the ninth week after the building was closed. The proposed method shows the best performance while On-LSTM-FLR outperforms Off-LSTM for this week. In Figure 4a, the prediction made by Off-LSTM has peaks in the mornings. This clearly shows that Off-LSTM still tries to predict the load using the old concept which leads to the worst performance of the three methods.

4.3.3. Post-COVID-19 Phase

For post-COVID-19 phase, the building is re-open and its electric load increases. However, the load pattern is different from that of the pre-COVID-19 phase and that of the COVID-19 phase. The test results for two weeks during the post-COVID-19 phase are presented.

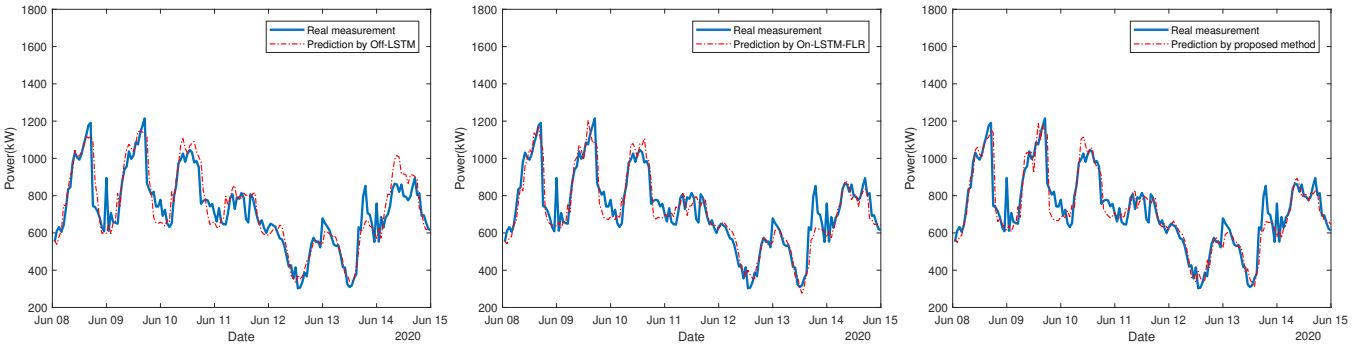
The results for the week from 5.25.2020 to 5.31.2020 is shown in Figure 7. This is the immediate week after the building was re-opened and corresponds to the second concept change. The proposed method shows the lowest MAE among the three methods. Off-LSTM has the largest MAE. All three methods fail to predict the load peak in the afternoon of May 25. Furthermore, they fail to predict the load peak at 23:00 of May 25 and 26. These new patterns have never been seen before.

The results for the week from 6.8.2020 to 6.14.2020 is shown



(a) Off-LSTM, MAE = 76.2 kW (b) On-LSTM-FLR, MAE = 69.1 kW (c) Proposed method, MAE = 67.5 kW

Figure 7. Results for post-COVID-19 phase, week from May 25 2020 to May 31 2020



(a) Off-LSTM, MAE = 59.1 kW (b) On-LSTM-FLR, MAE = 53.0 kW (c) Proposed method, MAE = 51.5 kW

Figure 8. Results for post-COVID-19 phase, week from Jun 8 2020 to Jun 14 2020

in Figure 8. The proposed method shows the lowest MAE among the three methods.

5. CONCLUSION

This paper proposed an online LSTM-based method for predicting building electric load during concept changes such as COVID-19. The proposed method uses a model ensemble maintaining three learners with different learning rates. The learning rates are adjusted online to find the best learning speed for the concept change. Experiment results show that the proposed method can quickly adapt the model to the concept changes during COVID-19 and reduce the prediction errors.

ACKNOWLEDGMENT

This work was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000666. The views and opinions of authors expressed herein do not necessarily state or reflect those of the US Government or any agency thereof.

The authors thank Nicola Peill-Moelter, Director Sustainability Innovation, VMware, Inc, for her support for this work.

REFERENCES

Beil, I., Hiskens, I., & Backhaus, S. (2016). Frequency regulation from commercial building hvac demand response. *Proceedings of the IEEE*, *104*(4), 745-757.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, *46*(4), 1–37.

Goddard, G., Klose, J., & Backhaus, S. (2014). Model development and identification for fast demand response in commercial hvac systems. *IEEE Transactions on Smart Grid*, *5*(4), 2084-2092.

Guan, X., Xu, Z., & Jia, Q. (2010). Energy-efficient buildings facilitated by microgrid. *IEEE Transactions on Smart Grid*, *1*(3), 243-252.

Guo, T., Xu, Z., Yao, X., Chen, H., Aberer, K., & Funaya, K. (2016). Robust online time series prediction with recurrent neural networks. In *2016 ieee international conference on data science and advanced analytics (dsaa)* (p. 816-825).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Kolter, J. Z., & Maloof, M. A. (2007). Dynamic weighted majority: An ensemble method for drifting concepts.

Journal of Machine Learning Research, *8*(Dec), 2755–2790.

Li, X., & Wen, J. (2014). Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews*, *37*, 517–537.

Mai, W., Chung, C., Wu, T., & Huang, H. (2014). Electric load forecasting for large office building based on radial basis function neural network. In *2014 ieee pes general meeting— conference & exposition* (pp. 1–5).

Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using deep neural networks. In *Iecon 2016-42nd annual conference of the ieee industrial electronics society* (pp. 7046–7051).

Saurav, S., Malhotra, P., TV, V., Gugulothu, N., Vig, L., Agarwal, P., & Shroff, G. (2018). Online anomaly detection with concept drift adaptation using recurrent neural networks. In *Proceedings of the acm india joint international conference on data science and management of data* (pp. 78–87).

Tu, H., Du, Y., Yu, H., Dubey, A., Lukic, S., & Karsai, G. (2020). Resilient information architecture platform for the smart grid: A novel open-source platform for microgrid control. *IEEE Transactions on Industrial Electronics*, *67*(11), 9393-9404.

Wang, S., & Xu, X. (2006). Simplified building model for transient thermal performance estimation using ga-based parameter identification. *International journal of thermal sciences*, *45*(4), 419–432.

Wang, Y., Wang, B., Chu, C.-C., Pota, H., & Gadh, R. (2016). Energy management for a commercial building microgrid with stationary and mobile battery storage. *Energy and Buildings*, *116*, 141–150.

Yang, H., Pan, Z., & Tao, Q. (2017). Robust and adaptive online time series prediction with long short-term memory. *Computational intelligence and neuroscience*, *2017*.

Yun, K., Luck, R., Mago, P. J., & Cho, H. (2012). Building hourly thermal load prediction using an indexed arx model. *Energy and Buildings*, *54*, 225–233.