Residential Energy Consumption Prediction Based on Encoder-Decoder LSTM

Junni Su, Lide Zhou, Fengchao Chen(✉), and Zejian Qiu

Dongguan Power Supply Bureau, Guangdong Power Grid Corporation, Dongguan 523000, Guangdong, China
csgcfc@126.com

Abstract. Accurate forecast of load profile is of great benefit to electricity dispatch and power grid management. Recently, the widespread of smart meters enable the power system to collect fine-grained data from massive users. Also, the development of deep learning techniques allow the load forecasting to have better performance. However, the hyperparameter tuning in neural networks is a laborious but ineluctable part to achieve higher accuracy. Combing with huge information concealed in the fine-grained data, data mining is a significant process to accelerate hyperparameter tuning. In this paper, we first explore the metadata to help filter the data, compare the performances with different input and output by varying granularities, and evaluate predictability on various aggregation levels. Numerical studies suggest that on filtered data, accuracy has a higher correlations with predictability, and granularity of 1 h is the most appropriate.

Keywords: Load Forecast · LSTM · encoder-decoder

1 Introduction

The deployment of smart meters and other intelligent terminals enables the frequent interactions between power companies and customers. Collecting and analyzing the load profile from user ends facilitate the load monitoring [16], appliance control [17], power grid management [11], and electricity price [22]. As a time series dataset, load profiles mainly refer to the users' electricity demand over time. And intuitively, choosing short-term forecasting generally has less training time and have quick adaptability. However, the procedure of increasing forecasting accuracy should follow the framework of data mining and include data preprocessing, feature selection, model selection and evaluation. Therefore, data libraries and supporting dataset (metadata) are recommended to help clean the main dataset. But in recent years, deep learning techniques such as recurrent neural network (RNN) are widely applied to tackle time series forecast problems instead of traditional times series models such as ARIMA model.

This work was supported in part by the science and technology study of China Southern Power Grid Cooperation (No. GDKJXM20200475).

H. Yang et al. (Eds.): SmartGIFT 2022, LNICST 483, pp. 319–330, 2023.
https://doi.org/10.1007/978-3-031-31733-0_27

1.1 Related Work

Short term load forecast is a challenging task due to the data uncertainty [14] and nonsmooth and nonlinear behaviors in load profiles [1]. Ding *et al.* [4] has shown that neural network based predictions outperforms time series models and the model design affects the generality abilities. Aowabin Rahman *et al.* [10] has demonstrated that RNN model has a high performance on forecasting tasks but fine tuning issues cannot be avoiding when training the model. In deep learning theory, we can generally increase the width to make the loss landscape smooth [8,13] and the depth to strengthen the representation power [2]. This insight is feasible in RNN model and deeper RNN model like including encoder-decoder framework is able to improve the performance [3]. Evaluation steps previously focus on measuring the distance between target and predictions. Instead, the evaluation of the input are generally ignored. Hence, predictability is adopted to help answer the question "when the input is most predictable" [18].

1.2 Our Research Contribution

In this report, we conduct the demand forecasting on Pecan Street Dataset. The principal contributions are:

– *Understanding the Dataset*: Any dataset must have an underlying background. By analyzing the metadata and weather data in Sect. 2, we filter the whole data into a certain group with same location and same building type as well as focus on a certain period of time without large weather changes.
– *Model Selection*: Instead of applying LSTM model directly, we discuss the effect of the input and output length on the prediction accuracy evaluated by MAPE and MPE. Also, we introduce encoder-decoder LSTM framework to increase the representation power and capacity of the neural network.
– *Predictability and Prediction Performance*: We define the predictability of a time series using the entropy of time and load level. After examining the predictability of different granularity and aggregation level, we find that small aggregation level has a positive effect on predicting the time series.

1.3 Paper Organization

The remainder of the paper is organized as follows. Section 2 presents the data overview and explores the metadata to filter users in the main dataset. Section 3 introduces the encoder-decoder LSTM framework and predictability definition. Section 4 presents the experiments design and corresponding results. Section 5 concludes and discuss future directions.

2 Understanding the Dataset

Pecan Street Dataport [12] is the world's largest residential energy and water research database. Data is organized into schemas by type of data (electricity, water, gas, static data) rather than by geographic location. Residential data is available as Time-series datasets (1-second energy, 1-minute energy, ISO data, water data, and natural gas data). Non-time series datasets (audits, surveys, and others) are available through Direct Database Access. Each home is linked to a unique Data ID.

To ensure the dataset is valuable and applicable, it is necessary to consider the background of dataset and explore the topic we want. Ideally, each user tend to be independent with others when the source is unique. But External factors such as the weather and economic currents inevitably influence users' behaviors. Hence, the data mining methods have to be time efficient and compatible with feature correlations.

2.1 Dataset Overview

Provided Pecan Street Dataset has three columns - "localminute", "Dataid" and "use". The "localminute" ranges from Jan 01, 2016 to Feb 22, 2016 recorded at 1 min level. In total, we have 339 unique data ID. After examining the missing value of each user, we found that data of user "1718", "2510", "3719", "7017", "7731" is not complete from its beginning time to its end time. Moreover, the "use" column represents the load of a certain user at 1 min level. Clearly, the type of data is a time series and the straightforward prediction on this series can be easily achieved by deep learning models. But the challenge is to seek for strategic or practical meanings of this dataset. What feasible direction can be discussed and utilized in companies?

2.2 Exploring the Metadata

For "localminute", the time can represent the local weather or even economic and political environments. Compared to the weather data, the latter one is hard to be quantified and likely to increase the model complexity tremendously if included. For now, we only consider the weather data in Austin (provided by [5]) and start from a short term forecasting to control the weather in a uniform level.

For "dataid", the identity of a user includes the information of location and the type, total area, construction year of user's building. Among all 339 users, we follow the rule of selecting the majority to reduce the variance. After examining the distribution and correlation of selected variables, we found that the relationship between them are not straight to be discovered. Therefore, we only select users from Austin with building type "Single-Family Home 001 (Master)".

In contrast to the data downloaded from the website, the "use" column is clearly an artificially modified feature. Except "localminute" and "Dataid", the raw dataset has 77 features. Searching through the metadata description, we found that the use column is close to the "grid" column, which presents measuring power drawn from or fed to the electrical grid. Though metadata provides the information about a large set of appliance that appears in users' houses, the sum of all appliance in table is not equal to the "grid" value. In other words, not all appliances in each user's home is monitored or recorded. Hence, only the "grid" column, the total amount of data can be valuable to this project. Due to incompleteness of appliance information, we cannot decompose it into the sum of all appliances in the home.

In fact, the information in the metadata is much more than what we mentioned before. For example, we may be interested in exploring whether the house total area and house construction year may have an effect on the load. However, Fig. 1 and 2 suggested that the correlation is unclear.

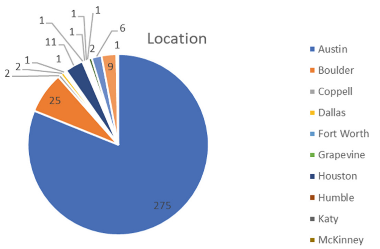


Fig. 1. Location

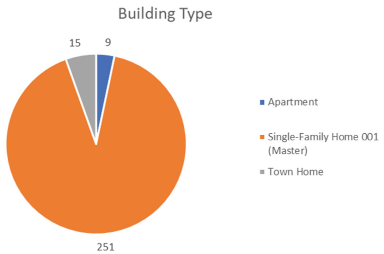


Fig. 2. Building Type

3 Model Selection and Evaluation

With the support of metadata and weather data, it is suggested to start from a short term load forecasting on selected users (218 in total). In this section, we will specify the goal of this paper and discuss how to select and evaluate the model. Empirically, the goal of predicting resident electric load varies by the project context. In other words, the input and output of a prediction depends on the feature selection. Since prior knowledge about the length of input and output is unknown, we begin by thinking about what prediction is meaningful.

3.1 Defining the Input and Output

If the input and output are both time series, then each represents a time interval. Equivalently, the question becomes given a certain time interval $\hat{\mathcal{T}}$ with length l_1, what length l_2 of time interval $\mathcal{T}$ before $\hat{\mathcal{T}}$ is most likely to be considered. With sufficient computation power and model capacity, the values of l_1 and l_2 are negligible since the prediction accuracy can be improved through trial and error, or fine-tuning. From the perspective of a electric power company, 1 min exact prediction with high accuracy are not necessary and in fact hard to achieve. To relax the strict requirements, we adopt the following steps:

1. sequence with small granularity → sequence with small granularity
2. sequence with small granularity → scalar with large granularity
3. sequence with small granularity → sequence with large granularity

Typically, using the predictions of “sequence to sequence” is the most straightforward and intuitive way. But the prediction accuracy is generally not satisfying. With the above steps, one may discover the prediction pattern with variation of accuracy. With high accuracy, it further provides insights for demand forecasting and user behavior analysis.

3.2 Modelling Method

The recent researches show a tendency to choose RNN models, especially LSTM model in the prediction task. Traditional time series methods and machine learning methods are fast to train, but the performance and the statistical properties (e.g., stationary sequence) impede their popularity in practice. Hence, in this section, we focus on how to improve the performance of LSTM model by modifying its structures.

In terms of the performance of a deep neural network, there is current belief that increasing the depth of a neural network will increase the representation power of and increasing the width of a layer will increase the smoothness of neural network landscape [2,8,13]. Note that one dimension of the input in LSTM models includes the parameter the number of features selected for a single user. Therefore, to increase the width, we introduce more features into the LSTM model such as the difference sequence or second difference sequence. Also, we increase the layer of the LSTM model. To increase the depth, modifying hidden layer in a single LSTM model cell is not the focus of this project. Hence, we propose to connect LSTM cells with other neural networks including encoder-decoder framework to form a hybrid model (encoder-decoder LSTM). The encoder-decoder LSTM consists of two LSTMs. The first LSTM, or the encoder, processes an input sequence and generates an encoded state. The encoded state summarizes the information in the input sequence. The second LSTM, or the decoder, uses the encoded state to produce an output sequence. Note that the input and output sequences can have different lengths [9]. We will build a encoder-decoder LSTM using PyTorch to make sequence-to-sequence predictions for time series data.

3.3 Evaluation Metric Extension

For sequence prediction, the evaluation metric is ususally Mean Absolute Percentage Error (MAPE) or Symmetric Mean Absolute Percentage Error (SMAPE). For scalar prediction, the evaluation metric is Mean Sqaured Error (MSE) or Mean Absolute Error (MAE). All the previous evaluation methods are aimed at measuring the performance of the output. In this section, we propose to extend the evaluation metric by adding predictability as defined in [18], a metric to evaluate the quality of input.

Intuitively, a model is most predictable over time if it follows an underlying function or a deterministic model while it is most unpredictable over time if it has largest uncertainty with respect to the load level. Therefore, the predictability are negatively correlated with conditional uncertainty of load over time.

Now, we present the calculation in detail. Given a frequency table with c rows of load level and t columns of time slots, every entry N_{ij} denotes the frequency of a sequence lying in the load level within certain time slot. Then the entropy of time $H(T)$ and load level $H(C)$ is defined as follows:

$$\begin{aligned} Z &= \sum_{1}^{c}\sum_{1}^{t} N_{ij} \\ \mathbb{P}(C_i) &= \sum_{1}^{t} N_{ij}/Z \\ \mathbb{P}(T_i) &= \sum_{1}^{c} N_{ij}/Z \\ H(C) &= -\sum_{1}^{c} \mathbb{P}(C_i)\log(\mathbb{P}(C_i)) \\ H(T) &= -\sum_{1}^{t} \mathbb{P}(T_i)\log(\mathbb{P}(T_i)) \end{aligned}$$

The property of entropy gives the result that $H(C)$ and $H(T)$ reaches the maximum $\log(c)$ and $\log(t)$ when each sum of row and sum of column are the same. Then, by the definition, the formula of calculating predictability $\mathcal{P}$ is given by:

$$\begin{aligned} \mathcal{P} &= 1 - \frac{H(CT) - H(T)}{\log c} = 1 - \frac{H(C|T)}{\log c} \\ H(CT) &= -\sum_{1}^{c}\sum_{1}^{t} \frac{N_{ij}}{Z}\log(\frac{N_{ij}}{Z}) \end{aligned}$$

Note that after dividing by $\log c$, we can actually scale the predictability into $[0, 1]$. Before we interpret the predictability, we first give the definition of constancy $\mathcal{C}$ and contingency $\mathcal{M}$ as below:

$$\begin{aligned} \mathcal{C} &= 1 - \frac{H(C)}{\log c} \\ \mathcal{M} &= \frac{I(CT)}{\log c} = \frac{H(C) + H(T) - H(CT)}{\log c} \end{aligned}$$

Now, from all the formula above, the following observations naturally follow:

1. $\mathcal{P} = \mathcal{C} + \mathcal{M}$: the predictability can be decomposed by the sum of constancy $\mathcal{C}$ and $\mathcal{M}$.
2. The contingency $\mathcal{M}$ measures the interaction between the load level and the time. The contingency reaches the maximal $\log(\min\{c, t\})$ means the load level completely depends on the time. In the meantime, the predictability equals to 0 if $c < t$, which is consistent with our intuition.

3. As in each time slot, there are same number of records, then $H(T) = \log(t)$ naturally follows in any situations.
4. The predictability actually measures how much does $H(C)$ lie in the whole load level entropy space. We illustrate this observation in the Fig. 3. Therefore, the predictability is a well defined metric.

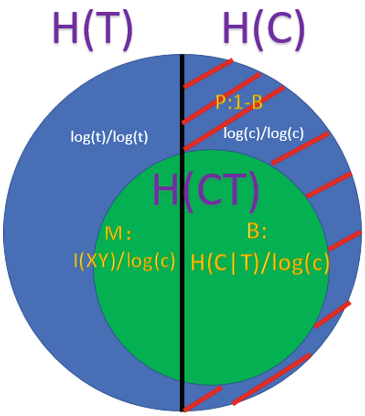


Fig. 3. The whole space denotes the maximum of $H(CT)$, which is reached when the mutual information is 0 and both $H(C)$ and $H(T)$ reach their maixmums respectively. According to the figure, high predictability implies the area B is small. Also, area $P - M$ denotes the constancy.

Having investigated the property of predictability, we now design the experiment setup to verify its effect. There are mainly two key parameters of predictability - granularity and aggregation level. Following Sect. 3.1, we conform the granularity steps and add one more prediction group. In previous method, we aim at predicting the results of each users while in the additional group, we first take an increasing sequence of subsets, sum up all inputs and outputs for each subset and evaluate their performance.

4 Experiment and Results

4.1 Experiment Design

In Sect. 3.1 and 3.3, the basic steps are outlined without specifying the parameters. The first part will not include the aggregation level and encoder-decoder LSTM. The purpose of this part is mainly to explore how LSTM performs on different input-output pairs. For granularity less than 3 h, we select time interval from Jan. 1^{st} to Jan. 2^{nd}. For granularity 3 h, we need to enlarge the time period since the input is not sufficient (Table 1).

Table 1. Experiment Setups

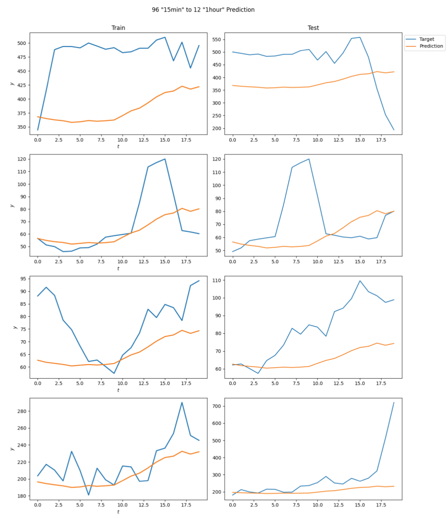
Input Length	Input Granularity	Output Length	Output Granularity
1440	1 min	60	1 min
1440	1 min	1	1 h
96	15 min	60	1 min
96	15 min	1	1 h
48	30 min	6	30 min
48	30 min	3	1 h
48	30 min	2	30 min
48	30 min	1	1 h
24	1 h	3	1 h
24	1 h	1	1 h
24	3 h	4	3 h
24	3 h	1	3 h

According to the results, we select best two input-output pairs and try them with different aggregation level and encoder-decoder LSTM model. In particular, the aggregation levels corresponds to the 2% to 100% of the all Austin Single Family users with stride 2%. The order of the users are randomized rather than sorted by "Dataid".

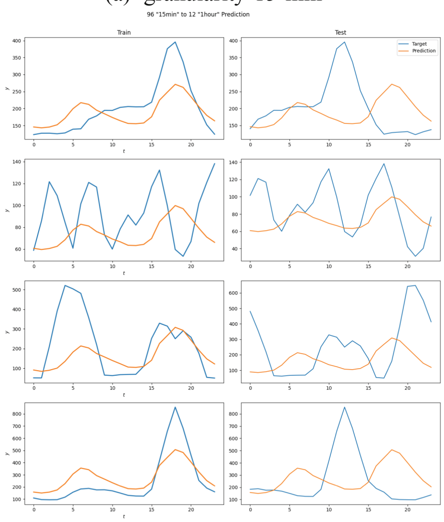
4.2 Results and Analysis

After trying the different input and output length pairs, we conclude that the input granularity 1 min and 3 h are not suitable for prediction, as shown in Fig. 4. The former one leads to a lengthy input and it is hard to train even after fine tuning and batch normalization. The latter one leads to a large data variance. If we treat each user as a single batch, then the variance among users can be large and a uniform learning rate with training techinques such as Xavier Normal, Stochastic Gradient Descent and weight decay still cannot achieve satisfying results. The prediction are almost same for different users. Therefore, we prefer using granularity 15 min and 30 min as our inputs, and the output can be a sequence of data with granularity 1 h or even 3 h. Figure 5 shows the results with input of different granularities. The results also suggest that LSTM generally performs well for a sequence-to-sequence prediction and there is no implication suggesting a sequence-to-scalar prediction is much better than a sequence-to-sequence prediction. Figure 5 present the value of predictability, constancy and contigency of different aggregation levels. Predictability sharply decreases when the aggregating level increases and keep at a low level later on. According to the

(a) granularity 15 min



(b) granularity 30 min

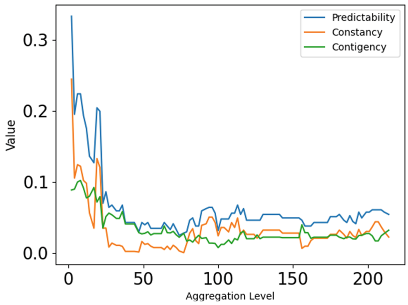


(c) granularity 1 hour

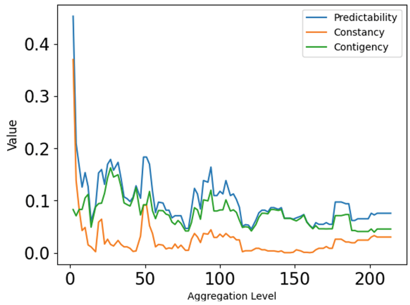
(d) granularity 3 hour

Fig. 4. Sequence-to-sequence predictions of different granularities

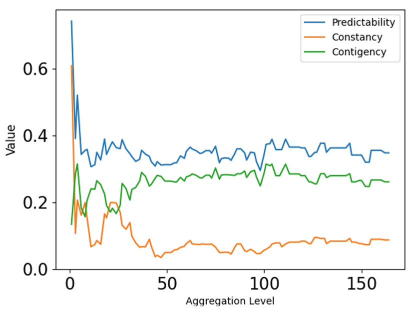
results, we run the proposed encoder-decoder LSTM model with granularity of 15 min and aggregation level 10 and 21 to assess how the aggregation level affect the prediction results. Figure 6 shows that increasing aggregation level may be of little benefit for prediction accuracy, which also implies that predictability has a high positive correlation with the performance of prediction model.



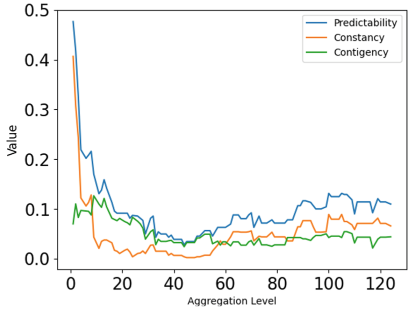
(a) granularity 15min predictability



(b) granularity 30min predictability

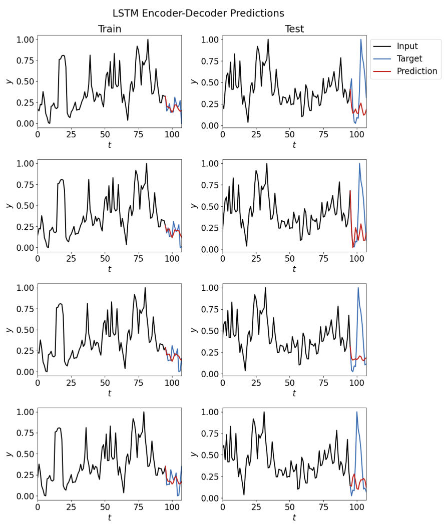


(c) granularity 1hour predictability

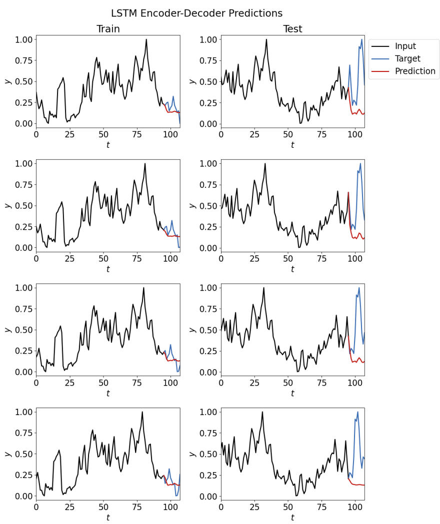


(d) granularity 3hour predictability

Fig. 5. Predictability (constancy and contingency) of different granularities



(a) prediction with aggregation level 10



(b) prediction with aggregation level 21

Fig. 6. Sequence-to-sequence prediction of different aggregation levels

5 Conclusions

Data mining techniques can extract technical framework in smart grid topics to facilitate generation dispatches and electricity management. In this project, we seek to forecast load based on Pecan Street individual load dataset. We include metadata and weather data to explore apt preprocessing procedure. Due to the limited size of dataset, we only focus on a certain group and select short term to alleviate the effect of weather. Compared with standard LSTM model, encoder-decoder LSTM model achieves better performance in the sense of MSE and MAE. The result of granularity selection shows that both small and large granularity are not suitable for prediction. After including predictability of data, the results suggest that aggregating data from a subset of users that has higher predictability provides more accurate predictions. Our experiments also have some limitations. From the supply end, filtering may lead to large bias of prediction. In other words, the generalization performance should be discussed and improved. Also, including random forest and AdaBoost may help reduce the performance variance. The encoder-decoder LSTM model is a very powerful learning framework, which can also be applied to many other learning tasks in the electricity sector, e.g., user profiling [15,19], learning-aided storage control [20,21], LMP prediction [6,7], etc.

References

1. Amjady, N., Keynia, F., Zareipour, H.: Short-term load forecast of microgrids by a new bilevel prediction strategy. IEEE Trans. Smart Grid **1**(3), 286–294 (2010)
2. Chatziafratis, V., Nagarajan, S.G., Panageas, I., Wang, X.: Depth-width trade-offs for Relu networks via Sharkovsky's theorem. arXiv preprint arXiv:1912.04378 (2019)
3. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
4. Ding, N., Benoit, C., Foggia, G., Besanger, Y., Wurtz, F.: Neural network-based model design for short-term load forecast in distribution systems. IEEE Trans. Power Syst. **31**(1), 72–81 (2015)
5. Apple Inc.: Dark sky. Website (2022). https://darksky.net/forecast/40.7127,-74.0059/us12/en
6. Jiang, G., Cui, J., Wu, C.: Role of prediction in data-driven pricing schemes. In: 2021 5th International Conference on Smart Grid and Smart Cities (ICSGSC), pp. 125–129. IEEE (2021)
7. Jiang, W.J., Cao, S., Wu, C.: LMP prediction with incomplete information. In: 2022 IEEE Power & Energy Society General Meeting, pp. 1–5. IEEE (2022)
8. Li, H., Xu, Z., Taylor, G., Studer, C., Goldstein, T.: Visualizing the loss landscape of neural nets. Adv. Neural Inf. Process. Syst. **31** (2018)
9. lkulowski: Building a LSTM encoder-decoder using PyTorch to make sequence-to-sequence predictions. https://github.com/lkulowski/LSTM_encoder_decoder. Accessed 20 Nov 2020
10. Rahman, A., Srikumar, V., Smith, A.D.: Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl. Energy **212**, 372–385 (2018)

11. Raza, M.Q., Khosravi, A.: A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renew. Sustain. Energy Rev. **50**, 1352–1372 (2015)
12. Street, P.: Pecan street dataport. Website (2016). https://dataport.pecanstreet.org
13. Sun, R., Li, D., Liang, S., Ding, T., Srikant, R.: The global landscape of neural networks: an overview. IEEE Signal Process. Mag. **37**(5), 95–108 (2020)
14. Tian, C., Ma, J., Zhang, C., Zhan, P.: A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. Energies **11**(12), 3493 (2018)
15. Wang, H., Luo, X., Wu, C.: Differential privacy in consumer behavior analysis. In: 2021 IEEE Power & Energy Society General Meeting (PESGM), pp. 1–5. IEEE (2021)
16. Wang, H., Zhang, J., Lu, C., Wu, C.: Privacy preserving in non-intrusive load monitoring: a differential privacy perspective. IEEE Trans. Smart Grid **12**(3), 2529–2543 (2020)
17. Wang, Y., et al.: Tracking states of massive electrical appliances by lightweight metering and sequence decoding. In: Proceedings of the Sixth International Workshop on Knowledge Discovery from Sensor Data, pp. 34–42 (2012)
18. Wu, C., Tang, W., Poolla, K., Rajagopal, R.: Predictability, constancy and contingency in electric load profiles. In: 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 662–667 (2016). https://doi.org/10.1109/SmartGridComm.2016.7778837
19. Wu, C., Tang, W., Poolla, K., Rajagopal, R.: Predictability, constancy and contingency in electric load profiles. In: 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 662–667. IEEE (2016)
20. Wu, J., Lu, C., Wu, C.: Learning-aided framework for storage control facing renewable energy. IEEE Syst. J. (2022)
21. Wu, J., Wang, Z., Wu, C., Wang, K., Yu, Y.: A data-driven storage control framework for dynamic pricing. IEEE Trans. Smart Grid **12**(1), 737–750 (2020)
22. Yue, S., Chen, J., Gu, Y., Wu, C., Shi, Y.: Dual-pricing policy for controller-side strategies in demand side management. In: 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 357–362. IEEE (2011)