A Multi-Step Time-Series Clustering-Based Seq2Seq LSTM Learning for a Single Household Electricity Load Forecasting

Zaki Masood, Rahma Gantassi, Ardiansyah and Yonghoon Choi *

Department of Electrical Engineering, Chonnam National University, Gwangju 61186, Korea; 186204@jnu.ac.kr (Z.M.); rahmagantassi@gmail.com (R.G.); ardi@ejnu.net (A.)

* Correspondence: yh.choi@jnu.ac.kr

Abstract: The deep learning (DL) approaches in smart grid (SG) describes the possibility of shifting the energy industry into a modern era of reliable and sustainable energy networks. This paper proposes a time-series clustering framework with multi-step time-series sequence to sequence (Seq2Seq) long short-term memory (LSTM) load forecasting strategy for households. Specifically, we investigate a clustering-based Seq2Seq LSTM electricity load forecasting model to undertake an energy load forecasting problem, where information input to the model contains individual appliances and aggregate energy as historical data of households. The original dataset is preprocessed, and forwarded to a multi-step time-series learning model which reduces the training time and guarantees convergence for energy forecasting. Furthermore, simulation results show the accuracy performance of the proposed model by validation and testing cluster data, which shows a promising potential of the proposed predictive model.

Keywords: deep learning; energy management system; LSTM; load forecasting; smart grid; time-series prediction

1. Introduction

The smart grid (SG) refers to the electric grid as an intelligent network of generation, transmission, and distribution. In the SG network, one of the benefits for the consumer is to generate energy using renewable resources while sustaining consumption and production with demand. The SG technologies are made possible by two-way communication, which involves data processing and control systems. The smart meter placed near a proximity network connects consumers with the grid and transmits data using advanced metering infrastructure (AMI) [1–5]. The AMI data help to analyze the distribution network and provide bidirectional communications to the consumer [2], which consolidates pricing information and demand response (DR). The recent advancement in the SG with effective demand-side management (DSM) allows big data into substantial benefits for the utilities and the customers. As reported [6], a smart meter can transfer energy information usage every 15 min, thus every million meters can generate 96 million reads in a day. However, due to the machine learning (ML) limitations [7–11], there has been a significant increase in deep learning (DL) models to exhibit complex correlations from a large dataset with numerous formats and replacements of manual feature extraction [12]. As reported [13], the expansion of AMI and wide-area monitoring systems (WAMS) in the SG network has increased the essential of DL techniques to deal along with the massive data. Moreover, a DL-based energy forecasting model is proposed to learn the correlation among distinct consumption behaviors for short-term load forecasting (STLF) [14–16]. The growth of literature on STLF for the individual household has demanded more practical analysis in the past decade. Specifically, the most commonly used DL techniques to address STLF problem include recurrent neural network (RNN), long short-term memory (LSTM), gated recurrent unit (GRU), convolutional neural network (CNN), reinforcement learning, autoencoders...
(AEs), and restricted Boltzmann machines (RBM). For example in [17], two state-of-art deep neural networks (DNN) architectures, deep feed neural network (FNN) and deep-RNN have been proposed for the STLF. The RNN is suitable for time-series data, where the network’s output is fed back to the input. The major drawback of using RNNs is that it is likely to vanish and explode gradient problems. For SG applications, this problem is overcome with memory gated structures that are LSTM and GRU.

In the recent past, it has been shown that load clustering [18–20] and multi-step load forecasting [21–24] techniques have been used to uncover power usage patterns established on specific measures of power data. The electric load clustering determines the power consumption patterns for a sustainable SG energy network with the demand-side response (DSR). However, based on the cluster data, it allows utilities to improve their policy and infrastructure planning techniques. Recently, Izonin et al. [25] and Tkachenko [26] proposed a non-iterative geometric transformations model (GTM), a neural-like structure discussed for solving time-series prediction tasks. Considering time-series forecasting problems, Ribeiro et al. [27] proposed a self-adaptive decomposed heterogeneous shallow model, predicting commercial and industrial electricity prices with multi-step-ahead commercial and industrial electricity prices. The optimal DL models drive more accurate forecasting and outperform the results of shallow structures models due to the limitations. For example, for the energy consumption forecasting problem, an LSTM-RNN-based univariate model [28] has been proposed, which forecast short- and medium-term time horizon few days up to weeks. On the other hand [29], a hybrid multivariate model has been proposed which combines CNN and LSTM for short- and long-term residential electricity forecasting. An increasing number of studies have found that home energy management systems (HEMS) and battery energy management systems (BEMS) have redirected the interest from aggregate loads towards data disaggregation of individual household loads. A novel clustering, classification, and forecasting (CCF) method has been proposed [30], outperforming the conventional smart meter-based model (SMBM) of individual household loads. The research activities for electric load clustering have increased continuously, and DL-based techniques for big data have recently demonstrated forecasting techniques for the SG network. The present information published to date has a significant viewpoint for DL models of aggregated residential loads that thoroughly improves the household load profiles and DR in the SG network. However, individual household load forecasting problems still need significant improvements and are required to be investigated in detail. The ENERTALK dataset is the first publicly available Korean dataset on electricity load consumption of 22 households with a sampling rate of 15 Hz. In this paper, we used the publicly available ENERTALK dataset [31]. The main contributions of this paper are as follows:

1. This paper proposes deep learning-based multi-step time-series Seq2Seq LSTM framework for the electricity load forecasting.
2. This paper takes on a new vision which combines the Seq2Seq LSTM and clustering to improve the efficiency of the DR program and provides multistep lookback analysis of a single household.
3. Different from the aggregated residential load, in this paper, a multi-step time-series electric load clustering and forecasting for a single household is proposed, which deals load forecasting to a DR program for supply and demand control.

We believe that our work presented in this paper will promote information communication technology (ICT) and artificial intelligence (AI) energy in SG networks.

This paper is organized as follows: Section 2 gives a brief overview of the proposed DL-based multi-step time-series Seq2Seq LSTM system model. In Section 3, we analyze load clustering and propose a multi-step time-series forecasting. Simulation results are presented in Section 4, followed by the conclusion in Section 5.
2. System Model

Electricity load forecasting is a challenging task for electricity utilities due to households’ different energy patterns and load characteristics. Nevertheless, it improves the utilities’ operational cost and estimates the electricity supply and demand to their customers. Figure 1 shows a proposed system model for a single household with a multi-step time-series clustering-based energy load forecasting strategy. In the system model, we use the 2.1 M samples of household 01 from the ENERTALK dataset. In the first stage, incomplete data such as noise, duplicates, and imbalances are handled by cleansing and normalization. Specifically, the data input size is reduced by extracting meaningful information for the later clustering and forecasting stage. Next, electric load clustering is essential for adopting an appropriate clustering algorithm for household data, where \( N \) represents the total number of clusters. Thus, each cluster is trained by the Seq2Seq LSTM in the final stage for better load dispatch and energy transfer scheduling.

![Figure 1.](image)

**Figure 1.** The framework of the proposed multi-step time-series electricity clustering and load forecasting system model. Time-series data preprocessing of aggregate load and downsampling of a single household. Before electricity load forecasting, each multi-step cluster datum is fed into a multi-step time-series Seq2Seq LSTM learning model.

### Data Preprocessing

It is known that data preprocessing is one integral part of the DL, which affects the learning ability of the DL model. This dataset contains the electricity consumption of 22 households from 1 September 2016 to 30 April 2017. Figure 2 shows the 122 days of energy consumption pattern from household 01, 02, and 03, where each appliance can be differentiated from a particular periodicity pattern to another. The active power recorded for appliances: refrigerator, rice cooker, washing machine, water purifier, and television. Considering regional features, kimchi-refrigerator [32] is commonly used in Korea, and also impacts global energy consumption.

The available dataset is large in length, which was originally sampled at 15 Hz. Therefore, we downsampled the electricity load of 122 days of household at every single minute into a multi-step time-series household consumption load profile. As a result, the downsampled pattern of electricity consumption behavior can be observed for the total household load of all appliances of 122 days in Figure 3. Specifically, considerable attention has been taken in advance to the SG dataset before applying it to the clustering and forecasting stage. Firstly, a null hypothesis was test enacted, which confirms a sample comes from a normal distribution [33]. For the distribution and normality test, \( s \) and \( k \) represent the skew and kurtosis for the dataset, respectively. It has been observed that both
$s$ and $k$ hold greater than zero, which confers statistical distribution is moderately skewed. The electricity load data is different for each appliance of all 122 days that can form an overfitting model. This statistical overfitting error can be minimized by partitioning the available data. The proposed system model overcomes this problem by adopting multi-step time-series clustering, which partitions the load data into parallel training and test data.

![Figure 2](image1.png)

**Figure 2.** Periodic patterns of home appliance of household ID = ‘01’, ‘02’, and ‘03’. Energy consumption behavior of ON state for households’ appliances washing machine, kimchi-refrigerator, refrigerator, water purifier, television, rice cooker, microwave, and the total power watts.

![Figure 3](image2.png)

**Figure 3.** Time-series daily load profiles without clustering (122 days from 1 October 2016 to 31 January 2017).

### 3. Proposed Framework

#### 3.1. Multi-Step Time-Series Electric Load Clustering

The clustering aims to find patterns of the load curves, which decide shifts in the load demand, specifically the groups of sample curves of original data values and their derivatives. We use the K-means algorithm, which finds centroids and groups of sample curves based on the nearest centroid value. The multi-step time-series electricity load patterns of a single household are summarized in Figure 4. It shows cluster centroids and related sample groups, which is obtained iteratively to reduce the sum of the Euclidean distances. Furthermore, we evaluate the goodness of the number of clusters for our multi-step time-series clustering algorithm by using the Silhouette score.

The Silhouette score evaluates intercluster distance for each sample within a cluster and intracluster distances among all clusters. It can be observed that the K-means algorithm with optimal silhouette score has four different clusters groups. The cluster groups 1 and 4 show the hourly load curve when consumption is low for all the 122 days. The red load curves have high peaks and later keep a steady load curve the rest of the time.
This prediction is noteworthy when employees mostly stay at home, which also includes weekends and special breaks. Cluster 3 has a high peak, but the rise is not steady most of the time compared to cluster 2, which can be observed as the fit load curve for business and school.

![Figure 4. Load characteristics curve of multi-step time-series clusters with number of clusters N = 4.](image)

3.2. Forecast Multiload Profiles

The basic RNN model works well for the sequence of data, and if valuable information from the cluster is modified or neglected, it will reduce the model’s accuracy [34]. Additionally, derivation during the backpropagation produces the vanishing gradient problem. Therefore, to overcome this problem for our forecasting model, we select one of LSTM’s architecture. It is known that LSTM has an artificial RNN chain structure, and the basic architecture of an LSTM model is composed of an input gate, output gate, and a forget gate. The function of these gates is to regulate the information through the LSTM network. Recently, research has shown the LSTM encoder and decoder architecture commonly used for language translation, where it processes the sequence from one domain to another. Furthermore, encoder–decoder architectures are conditional autoregressive models, generating a sequence from one domain to another. Therefore, we use the Seq2Seq model, which deals with the sequence of data to forecast our multi-step time-series problem.

This paper adopts the Seq2Seq LSTM model for a more substantial analysis of our multi-step time-series load forecasting problem. The architecture of a Seq2Seq LSTM model is shown in Figure 5, where each rectangular block holds an LSTM cell. Furthermore, each LSTM cell contains a hidden state, \( h_t \), and cell state, \( c_t \), at timestep, \( t \). The architecture is mainly divided into three parts: the encoder that is the input to the model, the decoder, which is the model’s output, and the encoder state vector. It can be seen that the encoder part is stacked with LSTM cells, and each of the cells allow a single element from the sequence as input.

The mathematical equations for the model are represented from (1) to (3). \( x = \{x_1, x_2, \ldots, x_T\} \) and \( y = \{y_1, y_2, \ldots, y_{T'}\} \) are the input time sequence and targeted output sequence of the forecasting model, respectively. It is important to note that the input sequence length \( T \) may differ from the target sequence length \( T' \). \( h_t \) and \( h'_t \) represent the encoder (1) and decoder (2) state from each LSTM cell at time \( t \), respectively. Specifically, \( x_t \) is the historical time-series data input to the LSTM cell at each timestep \( t \). Furthermore, the LSTM cell forwards the collected information from each cluster for training to the next LSTM cell. The output produced by each LSTM cell at time \( t \) is represented by \( h_t \), which is the encoder vector, the final hidden state produced by the encoder. The vector forwards the information for all input elements to the decoder for the predictions. \( y_t \) represents the target output sequence at a timestep \( t \). The LSTM encoder finds a conditional probability (3) by obtaining a hidden state vector \( v \) from the input sequence \( x \). The hidden state vector for the LSTM decoder learns conditional probability [35] for sequences of \( y \). Therefore,
by utilizing two different LSTM input and output sequences, learning is improved at the minimal cost of increased computational cost. Specifically, it benefits the network model to learn multi-step time-series simultaneously. The importance of the Seq2Seq model is that it can outline sequences of diverse time-series clustering lengths to each other. Next, the proposed framework forecast the load data for each cluster based on the train and test subsets.

\[
\begin{align*}
    h_t &= \text{LSTM}_{\text{encoder}}(x_t, h_{t-1}), \\
    h'_t &= \text{LSTM}_{\text{decoder}}(y_t, h'_{t-1}), \\
    p(y_1, y_2, \ldots, y_T|x_1, x_2, \ldots, x_T) &= \prod_{t=1}^{T} p(y_t|v, y_1, y_2, \ldots, y_{t-1}).
\end{align*}
\]

\[h_t = \text{LSTM}_{\text{encoder}}(x_t, h_{t-1}), \quad h'_t = \text{LSTM}_{\text{decoder}}(y_t, h'_{t-1}),\]

\[p(y_1, y_2, \ldots, y_T|x_1, x_2, \ldots, x_T)=\prod_{t=1}^{T} p(y_t|v, y_1, y_2, \ldots, y_{t-1}).\]

**Figure 5.** A Seq2Seq LSTM network model for the time-series load forecasting. At each timestep, the encoder takes one series of data \(x_t\) at time \(t\), and its previous state \(h_{t-1}\) and produces an output vector \(h_t\) state and cell state \(C_t\). The next decoder generates an output sequence \(y_t\), at each step taking at time \(t\), the previous state, and a weighted combination of all the encoder outputs (i.e., encoder state vector).

### 4. Numerical Analysis

Finally, the evaluation results are obtained from the proposed multistep time-series learning model. Table 1 shows the experiment parameters settings for all scenarios. We used LSTM, RNN, GRU, BiLSTM, and our proposed multi-step time-series Seq2Seq LSTM learning in these scenarios. The tests produce with Adam optimizer [36] on GPU enabled in a single NVidia to accelerate the computations. The platform adapts to develop the Seq2Seq LSTM learning in TensorFlow and Keras environment. In the simulation settings, we used timestamp as the time-series index. Each cluster datum is divided into training and testing subsets with a proportion of 67% and 33%, respectively. A further 25% of the training data are used to confirm the experiment for validation. Furthermore, we used different combinations of lookback periods, which find how many previous timesteps have been used to predict the subsequent timestep. For example, 60 lookback period shows the timestep at \(t - 60, t - 59, \ldots, t - 1\), and \(t\) has been used to predict the value at time \(t + 1\). Additionally, the dropout regularization [37] is set to 0.2, and MinMax scaling [38] is set between \(-1\) to \(1\) for the normalization of the dataset. Another hyperparameter is the batch size, set to 16, 64, and 128 combinations in the experiment. Moreover, for the result comparisons, we use the mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) metric evaluation.

**Figure 6** confirms that the cluster centroids and load curves are well apart and distinguished. It shows cluster mapping of 122 days electricity load data of a single household, where each data point is grouped into one of four different clusters. The closer data points are grouped as one cluster and represent similar load profiles. It can be seen that data points are close to the neighboring cluster, which confers the clustering is well molded.
Table 1. Parameters settings for the multi-step time-series learning model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecasting model</td>
<td>GRU, RNN, BiLSTM, Seq2Seq LSTM</td>
</tr>
<tr>
<td>Training dataset</td>
<td>67% of the total input</td>
</tr>
<tr>
<td>Testing dataset</td>
<td>33% of the total input</td>
</tr>
<tr>
<td>Validation split</td>
<td>25% of the train dataset</td>
</tr>
<tr>
<td>Minmax normalization</td>
<td>−1 to 1</td>
</tr>
<tr>
<td>Regularization</td>
<td>dropout 0.2 each layer</td>
</tr>
<tr>
<td>Number of sequence</td>
<td>64 units</td>
</tr>
<tr>
<td>Number of lookback</td>
<td>60, 120, and 180 periods</td>
</tr>
<tr>
<td>Number of maximal epochs</td>
<td>100</td>
</tr>
<tr>
<td>Optimization algorithm</td>
<td>Adam</td>
</tr>
<tr>
<td>Testing evaluation metrics</td>
<td>MAE, MAPE, RMSE</td>
</tr>
</tbody>
</table>

Figure 6. Results of time-series data points of K-means clustering.

For the comparative analysis, Table 2 shows the noteworthy results of the proposed forecasting model, and Figure 7 shows the performance comparison by using different lookback periods. The results of several learning approaches have been shown to improve the load forecasting with the same electricity load data. For the comparison, we used the existing implementation of state-of-the-art learning models, LSTM in [39], GRU in [40], and BiLSTM in [41]. However, we use the same parameter settings as in Table 1 for all learning models in this paper to make it comparable to the proposed multi-step time-series Seq2Seq LSTM learning. All learning models’ performance are observed using MAE, MAPE, and RMSE evaluation metrics. It is important to note that the results shown are adequate for the multi-step time-series Seq2Seq LSTM model with 60-, 120-, and 180-step periods, which are obtained after the clustering algorithm. It has also been observed that when the step size increased, the MAE, MAPE, and RMSE increased slightly, which indicates that the benefits of our multi-step time series exhibit more stable results with a large step size. The performance of our proposed multi-step Seq2Seq LSTM model is better than LSTM, GRU, RNN, and BiLSTM, and shows a stable improvement in Figure 7. Furthermore, it has been observed from the metrics performance that the simple RNN model’s performance is the worst out of all the tested models. However, the RNN model is not useful for our proposed multi-step time-series clustering-based electricity forecasting model. These results confirm the effectiveness of the proposed encoder–decoder LSTM approach for the multi-step time-series electricity load forecast compared to outputting a vector directly into learning model.
Table 2. Comparison of the value of MAE, MAPE, and RMSE associated to all scenarios with the proposed multi-step time-series Seq2Seq LSTM forecasting model, specifically shows three different lookback periods at 60, 120, and 180 timesteps.

<table>
<thead>
<tr>
<th>Forecasting Models</th>
<th>MAE</th>
<th>MAPE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM 60 timesteps</td>
<td>44.3</td>
<td>11.93</td>
<td>92.91</td>
</tr>
<tr>
<td>RNN 60 timesteps</td>
<td>55.7</td>
<td>17.99</td>
<td>102.37</td>
</tr>
<tr>
<td>GRU 60 timesteps</td>
<td>39.2</td>
<td>12.20</td>
<td>84.97</td>
</tr>
<tr>
<td>BiLSTM 60 timesteps</td>
<td>40.3</td>
<td>13.22</td>
<td>88.41</td>
</tr>
<tr>
<td><strong>Seq2Seq LSTM 60 timesteps (proposed)</strong></td>
<td><strong>35.1</strong></td>
<td><strong>10.93</strong></td>
<td><strong>82.75</strong></td>
</tr>
<tr>
<td>LSTM 120 timesteps</td>
<td>50.9</td>
<td>16.22</td>
<td>98.35</td>
</tr>
<tr>
<td>RNN 120 timesteps</td>
<td>61.2</td>
<td>20.10</td>
<td>109.28</td>
</tr>
<tr>
<td>GRU 120 timesteps</td>
<td>36.6</td>
<td>10.87</td>
<td>82.71</td>
</tr>
<tr>
<td>BiLSTM 120 timesteps</td>
<td>47.2</td>
<td>13.23</td>
<td>91.63</td>
</tr>
<tr>
<td><strong>Seq2Seq LSTM 120 timesteps (proposed)</strong></td>
<td><strong>46.5</strong></td>
<td><strong>12.22</strong></td>
<td><strong>86.50</strong></td>
</tr>
<tr>
<td>LSTM 180 timesteps</td>
<td>48.9</td>
<td>13.09</td>
<td>99.50</td>
</tr>
<tr>
<td>RNN 180 timesteps</td>
<td>67.3</td>
<td>22.47</td>
<td>113.92</td>
</tr>
<tr>
<td>GRU 180 timesteps</td>
<td>42.9</td>
<td>13.25</td>
<td>88.60</td>
</tr>
<tr>
<td>BiLSTM 180 timesteps</td>
<td>41.6</td>
<td>11.19</td>
<td>89.75</td>
</tr>
<tr>
<td><strong>Seq2Seq LSTM 180 timesteps (proposed)</strong></td>
<td><strong>38.5</strong></td>
<td><strong>13.32</strong></td>
<td><strong>88.65</strong></td>
</tr>
</tbody>
</table>

Figure 7. Example performance evaluation of our proposed model with other learning models with 60 timesteps.

This work finds different combinations of epochs and batch sizes to analyze the convergence Seq2Seq LSTM. The training performance of each architecture is evaluated by validation samples, including different numbers of epochs and batch sizes; to avoid overfitting, dropout has been used. Moreover, the model’s learning ability improved with the combination of time-series clustering and multi-step encoder–decoder sequences. Figure 8 shows the convergence of the loss function for the training and validation of the proposed model. It reveals that both training and validation loss decreases, and the proposed model obtained the convergence approximately after 20 epochs. In order to confirm that our proposed model also works well for the multi-step lookback periods, the training and testing have been carried out ranging from 1 to 200 lookback timesteps. Figure 9 shows the training, validation, and prediction curves for the 60-steps lookback periods. Furthermore, for the multi-step time-series forecasting, repeated executions of the proposed model have been performed by varying the number of lookback periods. For example, the actual load curves of a single household and its validation and predictive curves are shown in Figure 9 with the 60-steps forecasting. In correspondence with the testing data, the accuracy increases and the validation over the target data is achievable with the increase in size. Thus, the overall prediction of our proposed clustering-based...
Seq2Seq2 LSTM is the most suitable, and a very small portion of the data has a weak correlation.

![Figure 8](image)

**Figure 8.** Results of convergence of the loss function of the proposed multi-step time-series Seq2Seq LSTM learning. Train and validation loss with number of epochs and mean square error.

![Figure 9](image)

**Figure 9.** The actual load curves of household ID = ‘01’ of 122 days together with its validation and predictive curves from our multi-step time-series forecasting method.

5. Conclusions

This paper proposed a multi-step time-series clustering-based Seq2Seq LSTM learning model in order to forecast a single household’s electricity load. The proposed framework confirms multi-step time-series Seq2Seq-LSTM learning and compares it with other models (LSTM, RNN, GRU, and BiLSTM). The results demonstrate that the proposed model better adapts with the combination of clustering and 60-, 120-, and 180-step time series Seq2Seq load forecasting, which shows the best performance based on MAE, MAPE, and RMSE evaluation metrics. Furthermore, the simulation results showed that cluster-based multi-step time-series Seq2Seq LSTM learning significantly improves single household load forecasting. This confirms that the cluster based multi-step time-series learning is a reliable approach for the future load forecasting of households. Furthermore, the limitation of the univariate analysis can be extended to the multivariate with multi-step load forecasting in future work. This research will open further challenges for applying DL techniques to the SG network in the future.
Author Contributions: The research was carried out successfully with contribution from all authors. The main research idea and manuscript preparation were contributed by Z.M. and Y.C.; R.G. and A. contributed to the manuscript preparation and gave several suggestions from industrial perspectives. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1I1A3A01060631).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References


4. Masood, Z.; Ardiyansyah; Choi, Y. Energy-Efficient Optimal Power Allocation for SWIPT Based IoT-Enabled Smart Meter. Sensors 2021, 21, 7857. [CrossRef]


34. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys. D Nonlinear Phenom. 2020, 404, 132306. [CrossRef]