

Article

Asymmetric Loss Functions for Contract Capacity Optimization

Jun-Lin Lin ^{1,2,*} , Yiqing Zhang ^{1,3}, Kunhuang Zhu ^{1,3}, Binbin Chen ^{1,3} and Feng Zhang ^{1,3}

¹ Department of Information Management, Yuan Ze University, Taoyuan 32003, Taiwan; s1051770@mail.yzu.edu.tw (Y.Z.); kunhuang_zhu@163.com (K.Z.); s1051668@mail.yzu.edu.tw (B.C.); s1051768@mail.yzu.edu.tw (F.Z.)

² Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan 32003, Taiwan

³ School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China

* Correspondence: jun@saturn.yzu.edu.tw; Tel.: +886-3-463-8800 (ext. 2611)

Received: 16 May 2020; Accepted: 15 June 2020; Published: 16 June 2020



Abstract: For high-voltage and extra-high-voltage consumers, the electricity cost depends not only on the power consumed but also on the contract capacity. For the same amount of power consumed, the smaller the difference between the contract capacity and the power consumed, the smaller the electricity cost. Thus, predicting the future power demand for setting the contract capacity is of great economic interest. In the literature, most works predict the future power demand based on a symmetric loss function, such as mean squared error. However, the electricity pricing structure is asymmetric to the under- and overestimation of the actual power demand. In this work, we proposed several loss functions derived from the asymmetric electricity pricing structure. We experimented with the Long Short-Term Memory neural network with these loss functions using a real dataset from a large manufacturing company in the electronics industry in Taiwan. The results show that the proposed asymmetric loss functions outperform the commonly used symmetric loss function, with a saving on the electricity cost ranging from 0.88% to 2.42%.

Keywords: contract capacity; asymmetric loss functions; long short-term memory

1. Introduction

Predicting power demand is vital for power companies to plan the production of electricity. The overproduction of electricity not only increases the production cost and the electricity attrition rate but also accelerates the damage to the power equipment and the pollution to the environment. In contrast, the underproduction of electricity could result in energy rationing or even interruption of the power supply. Thus, the prediction of power demand has received much attention from both academics and practitioners in the electric power industry [1–6].

Various technologies have been proposed in the literature for forecasting power demand and optimizing energy consumption. For example, Building Energy Simulation, hybrid approaches and statistical-based tools were applied to building energy simulation models [7,8]. Furthermore, a smart grid co-simulation software platform between the energy management systems and a building energy model was proposed to decouple the control algorithms and the building [9]. In [10], a linear programming model and an energy system cost-optimization model were used to forecast the long-term power demand and CO₂ emissions. In [11], time-series forecasting models were adopted to forecast the long-term power demand in a scenario of energy transition from fossil fuels to carbon-free sources. In [12], an energy consumption scheduling device was used on the customer site to achieve autonomous demand response with the goal of controlling customer's flexible demand and optimizing the cost. For a recent review of state-of-the-art load forecasting techniques, please refer to [13].

The design of contract capacity in the electric power market is an effective means to control the power demand of high-voltage and extra-high-voltage consumers through demand-side management. A high-voltage or extra-high-voltage customer signs a contract with a power company to purchase a certain amount of electricity (referred to as “contract capacity”) for the next month. If the customer uses more electricity than the contract capacity during the next month, he/she needs to purchase the excess amount of electricity at a higher rate. However, if the customer uses less electricity than the contract capacity during the next month, he/she still has to pay for the remaining unused part of the contract capacity. Thus, it is advantageous for the customer to set the contract capacity near his/her actual power demand to avoid paying electricity at a higher rate. Additionally, because these high-voltage and extra-high-voltage consumers account for a large proportion of total electricity consumption, the power company can use their contract capacities to improve the prediction of the total power demand.

High-voltage and extra-high-voltage consumers need to predict their future power demand to decide their contract capacity. Previous work on predicting the future power demand is usually based on a symmetric loss function, such as mean squared error (MSE), to build the prediction models [5,6]. However, for high-voltage and extra-high-voltage consumers, their electricity pricing structure is asymmetric to the contract capacity. Specifically, underestimation (i.e., contract capacity < actual power demand) usually results in a higher per unit electricity cost than overestimation (i.e., contract capacity > actual power demand) does, for the same amount of difference between the actual power demand and the contract capacity. The motivation of this study is to adopt an asymmetric loss function to reflect the asymmetric pricing structure in the prediction model to minimize the electricity cost.

In this paper, we study the problem of determining the contract capacity to minimize the electricity cost for high-voltage and extra-high-voltage consumers. Since electricity cost usually accounts for a significant expense to high-voltage and extra-high-voltage consumers, even a small percentage of reduction in electricity expense could significantly reduce their overall operation cost, making them competitive in today’s market. In the literature, Long Short-Term Memory (LSTM) has shown to avoid the vanish gradient problem in the recurrent neural networks and yield excellent performance for time series prediction [6,14,15]. Thus, we use an LSTM neural network with an asymmetric loss function to build the prediction model for contract capacity, where the model is trained using the monthly electricity consumption data. The asymmetric loss function is derived from the pricing structure for high-voltage and extra-high-voltage consumers of Taiwan Power Company. Our performance study using a real dataset shows that the model based on the asymmetric loss function outperforms the model based on the commonly-used symmetric loss function, MSE.

The rest of this paper is organized as follows. Section 2 reviews previous work on electricity load forecasting and optimization of contract capacity. Section 3 formally defines the problem of determining the contract capacity to minimize the electricity cost. Section 4 proposes six asymmetric loss functions derived from the electricity pricing structure. Section 5 presents our performance study, and Section 6 concludes this paper and give directions for future research.

2. Related Work

The problem of determining the contract capacity to minimize the electricity cost is closely tied to the problem of predicting the future power demand. If we can predict the future power demand accurately, we can set the contract capacity accordingly to avoid paying electricity at a higher rate. Based on the scale of the forecast horizon, predicting the future power demand can be divided into four categories: very short-term, short-term, medium-term, and long-term. Very short-term and short-term load forecasting focuses on the prediction of hourly or daily load for one hour to four weeks ahead [2,4,14,16]. Medium-term forecasting aims to predict monthly load for 1 to 12 months ahead, and long-term load forecasting aims to predict yearly load for 1 to 20 years ahead [16]. Since contract capacity can be set on a month-to-month basis, medium-term load forecasting fits the requirement of determining the contract capacity.

In the literature, many forecasting methods have been proposed for electricity load forecasting [4,5]. For example, Seasonal Autoregressive Integrated Moving Average model (SARIMA) and feed-forward back-propagation neural networks were used to predict the power demand in the Turkish electricity market [17]. Among these forecasting methods, artificial neural networks, especially recurrent neural networks (RNNs), are the most widely used. RNN considers the ordering among data, making it very suitable for time series prediction. In [18], empirical mode decomposition on a sliding window was used to select features (including power demand features and weather-related features). Then, an Elman network (a variant of RNNs) was trained to predict the future power demand, where the weightings in the network were optimized by a population-based heuristic search algorithm.

However, vanilla RNN suffers from the vanishing gradient and the exploding gradient problems, making it unable to retain the useful memory about the data exhibited earlier. To mitigate this problem, LSTM adds more gates and links to an RNN node such that useful (or useless) memory about past data can be remembered (or forgotten). LSTM has been applied to many problems where time ordering of data is crucial, e.g., sequence translation [19], human activity recognition [20], hyperspectral image classification [21], and electricity load forecasting [6,14]. In [14], a prediction model for the electricity load of a day is constructed using the data from similar days in the past. First, the data from similar days are decomposed into several intrinsic mode functions (IMFs) using empirical mode decomposition. Then, an LSTM prediction model is built for each IMF, and finally, the load prediction is formed by combining the predictions from all these LSTM models. In [6], LSTM and the genetic algorithm are integrated for short-term load forecasting.

Because the electricity cost is minimized when the actual power demand is equal to the contract capacity, we can first predict the power demand for next month and then use the predicted value as the contract capacity for next month. However, most of the previous works on load forecasting aim to minimize the error between the predicted and the actual power demands without considering the pricing structure of electricity. Thus, the costs of underestimation and overestimation are symmetric. Consequently, they use a symmetric loss function, such as MSE, both to adjust the prediction model and to evaluate the performance of the predictions. However, as described in Section 1, the electricity pricing structure for high-voltage and extra-high-voltage consumers is asymmetric to the contract capacity. For the goal of minimizing the electricity cost, using an asymmetric loss function is more appropriate than using MSE.

There are two scenarios where an asymmetric loss function is often adopted. The first scenario is when the available dataset is imbalanced, e.g., in medical imaging applications [22]. The second scenario is when the loss of the underlying problem is asymmetric. One example is remaining useful life (RUL) estimation in Prognostics and Health Management (PHM), where underestimating the RUL of a component only results in the waste of replacing the component too earlier, while overestimating the RUL may cause detrimental effects to the machinery [23]. Other examples include the asymmetric loss on wind speed and power predictions [24,25] and oil price prediction [26]. In this paper, we focus on the problem of setting the contract capacity to minimize the cost. Most of the existing methods for load forecasting did not consider the electricity pricing structure, which is asymmetric to under- and overestimation. As a result, an unbiased function such as MSE was often adopted to train the model and to evaluate the prediction performance. Consequently, the same amount of underestimation and overestimation results in the same cost using MSE, which contradicts to the electricity pricing structure.

The problem of contract capacity optimization was studied in [27–30]. Given the monthly power demand for the past twelve months and the electricity pricing structure, the problem is to determine the contract capacity for each of these past twelve months such that the total electricity cost for these twelve months is minimized. This problem considers more facets in the electricity pricing structure than just the per unit electricity cost. For example, the electricity pricing structure also includes the expanding construction fee for those months of increasing the contract capacity. As a result, simply setting the contract capacity to the actual power demand does not necessarily minimize the total cost. This problem can be formulated as an optimization problem, and linear programming [27] and

metaheuristic algorithms [28,29] have been used to derive or search the optimal contract capacity for each month. Notably, this problem assumes that the real power demand is known, making it different from the problem studied in the current paper, where the real power demand is unknown, and the goal is to determine the contract capacity to minimize the electricity cost for future months.

3. Problem Formulation

This paper considers the problem of determining the next month's contract capacity for a high-voltage or extra-high-voltage consumer such that his/her electricity cost can be minimized. We adopt the electricity pricing structure of Taiwan Power Company and focus on the determination of peak contract capacity [27]. Let R be the basic per unit electricity cost, and \hat{x}_i and x_i denote the contract capacity and the actual power demand for month i of a consumer, respectively.

If the actual demand x_i is less than the contract capacity \hat{x}_i , then the customer has to pay a fixed capacity charge $R\hat{x}_i$. If the actual demand is greater than the contract capacity, then the excess demand within 10% of the contract capacity is charged at twice the basic rate R , and the excess demand over 10% of the contract capacity is charged at three times the basic rate. Thus, the customer's electricity cost for month i (denoted by C_i) can be calculated as follows.

$$C_i = \begin{cases} R\hat{x}_i & \text{if } x_i \leq \hat{x}_i \\ R(\hat{x}_i + 2(x_i - \hat{x}_i)) & \text{if } \hat{x}_i < x_i \leq 1.1\hat{x}_i \\ R(\hat{x}_i + 0.2\hat{x}_i + 3(x_i - 1.1\hat{x}_i)) & \text{if } 1.1\hat{x}_i < x_i \end{cases} \quad (1)$$

Ideally, if $\hat{x}_i = x_i$, then the consumer's electricity cost for month i equals Rx_i , which is the optimal electricity cost for the power demand x_i . If the contract capacity \hat{x}_i is an over- or underestimation of the actual demand x_i , then a penalty is imposed on the electricity cost. The penalty can be calculated as the electricity cost C_i minus the optimal electricity cost Rx_i . Let P_i^C denote the penalty on the customer's electricity cost for month i due to under- or overestimation of the actual demand. Because the basic per unit electricity cost R is a constant, we further divide $C_i - Rx_i$ by R to make the value of P_i^C independent of the value of R , as shown below.

$$P_i^C = \frac{C_i - Rx_i}{R} = \begin{cases} \hat{x}_i - x_i & \text{if } x_i \leq \hat{x}_i \\ x_i - \hat{x}_i & \text{if } \hat{x}_i < x_i \leq 1.1\hat{x}_i \\ 2x_i - 2.1\hat{x}_i & \text{if } 1.1\hat{x}_i < x_i \end{cases} \quad (2)$$

Based on Equation (2), we define the problem under study as follows.

Problem Definition. Given the monthly power demand up to month $i - 1$, exclusively, predict the contract capacity \hat{x}_i for month i such that the penalty P_i^C for month i is minimized.

Notably, when we need to decide the contract capacity for month i , the actual power demands for both months i and $i - 1$ are still unknown. Thus, x_i and x_{i-1} cannot be used to predict \hat{x}_i . In other words, this problem has a forecast horizon of 2.

To show the asymmetric pricing structure of electricity, we divide the electricity cost C_i by the actual demand x_i to yield the consumer's per unit electricity cost for month i (denoted as R_i), as follows.

$$R_i = \frac{C_i}{x_i} = \begin{cases} R\left(\frac{\hat{x}_i}{x_i}\right) & \text{if } x_i \leq \hat{x}_i \\ R\left(2 - \frac{\hat{x}_i}{x_i}\right) & \text{if } \hat{x}_i < x_i \leq 1.1\hat{x}_i \\ R\left(3 - \frac{2.1\hat{x}_i}{x_i}\right) & \text{if } 1.1\hat{x}_i < x_i \end{cases} \quad (3)$$

Then, the penalty (denoted as P_i^R) on the customer's per unit electricity cost for month i can be calculated as the per unit electricity cost R_i minus the basic per unit electricity cost R . Similar to Equation (2), we further divide $R_i - R$ by R to make the value of P_i^R independent of the value of

R (see Equation (4)). Notably, the if-conditions of Equation (4) are the same as the if-conditions of Equations (1)–(3), but are rephrased using $\frac{\hat{x}_i}{x_i}$ for ease of exposition.

$$P_i^R = \frac{R_i - R}{R} = \begin{cases} \frac{\hat{x}_i}{x_i} - 1 & \text{if } 1 \leq \frac{\hat{x}_i}{x_i} \\ 1 - \frac{\hat{x}_i}{x_i} & \text{if } \frac{1}{1.1} \leq \frac{\hat{x}_i}{x_i} < 1 \\ 2 - \frac{2.1\hat{x}_i}{x_i} & \text{if } \frac{\hat{x}_i}{x_i} < \frac{1}{1.1} \end{cases} \quad (4)$$

The solid blue line in Figure 1 shows the values of $\frac{\hat{x}_i}{x_i}$ versus the values of P_i^R . At $\frac{\hat{x}_i}{x_i} = 1$, both P_i^C and P_i^R are zero, indicating a perfect prediction. This blue line is asymmetric about the line $\frac{\hat{x}_i}{x_i} = 1$. For $1 \leq \frac{\hat{x}_i}{x_i}$ and $\frac{1}{1.1} \leq \frac{\hat{x}_i}{x_i} < 1$, the slopes of Equation (4) are 1 and -1 , respectively. As the value of $\frac{\hat{x}_i}{x_i}$ moves away from 1, the value of penalty P_i^R increases at the same speed for both overestimation (i.e., $1 \leq \frac{\hat{x}_i}{x_i}$) and underestimation (only in the range of $\frac{1}{1.1} \leq \frac{\hat{x}_i}{x_i} < 1$) of the actual demand. However, as the degree of underestimation aggravates (i.e., $\frac{\hat{x}_i}{x_i} < \frac{1}{1.1}$), the slope of Equation (4) changes to -2.1 from -1 , i.e., the speed of increasing penalty P_i^R is more than double of the original speed.

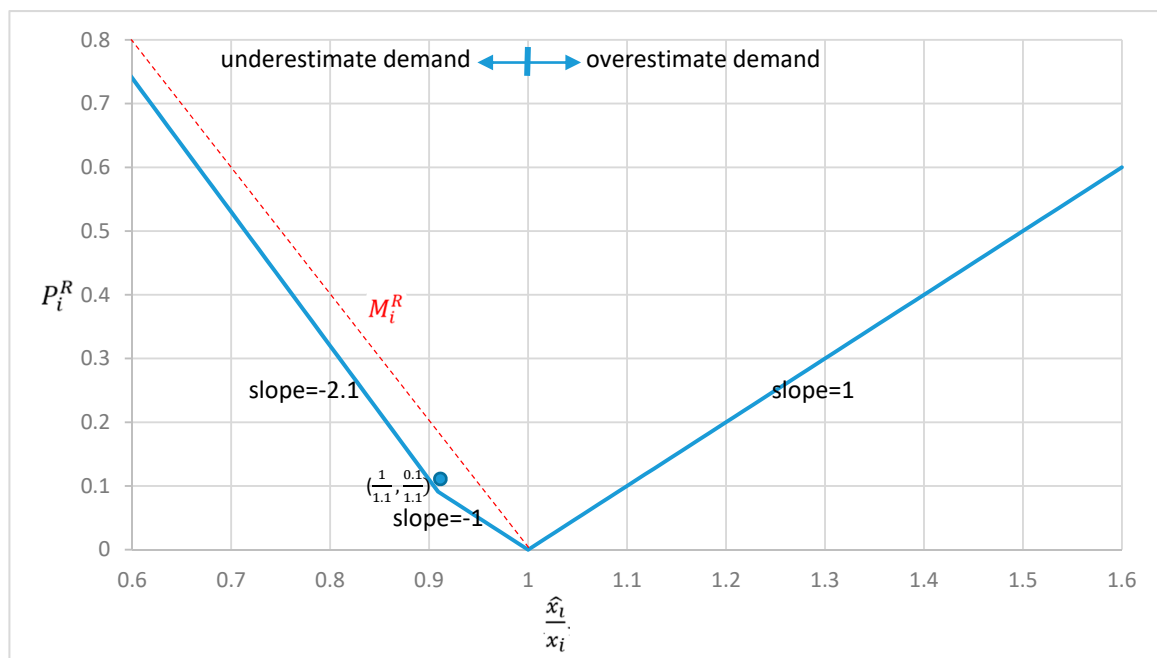


Figure 1. Penalty P_i^R v.s. the ratio of the contract capacity \hat{x}_i and the actual power demand x_i .

4. Proposed Loss Functions for Contract Capacity Prediction

In machine learning, a loss function can be used to adjust a prediction model to fit the training data. For example, an artificial neural network uses a loss function to calculate the difference between the predicted and actual values of the training data. It then back-propagates the difference to fine-tune the weightings of the links in the network. The most commonly used loss function is MSE, which is symmetric to under- and overestimation. In this study, we use the monthly contract capacity as the predicted values for the actual power demand, so MSE can be calculated as follows, where n denotes the number of instances in the training data.

$$MSE = \sum_{i=1}^n \frac{(\hat{x}_i - x_i)^2}{n} \quad (5)$$

In the rest of this section, we proposed several asymmetric loss functions that are derived from the electricity pricing structures described in Section 3. Later, in Section 5, we experiment with these loss functions and compare their performance against that of MSE.

As discussed in Section 3, the electricity pricing structure is asymmetric to under- and overestimation of the actual power demand. Thus, a loss function that takes into account this asymmetric pricing structure is more appropriate for the machine learning algorithm to optimize the electricity cost. In Section 3, we derived the penalty of electricity cost P_i^C in Equation (2) and the penalty of per unit electricity cost P_i^R in Equation (4). In Equations (6) and (7), we define two loss functions, L^C and L^R , by calculating the average of P_i^C and P_i^R over the training data, respectively.

$$L^C = \sum_{i=1}^n \frac{P_i^C}{n} \quad (6)$$

$$L^R = \sum_{i=1}^n \frac{P_i^R}{n} \quad (7)$$

Both L^C and L^R are zero when the predicted values equal the actual values. Furthermore, underestimation incurs more penalty to both L^C and L^R than overestimation does.

The MSE loss function has the effect of slowly increasing its value as the deviation between the predicted value and the actual value is small, but quickly increasing its value as the deviation gets large. We can modify Equations (6) and (7) as follows to achieve a similar effect.

$$L^{C2} = \sum_{i=1}^n \frac{(P_i^C)^2}{n} \quad (8)$$

$$L^{R2} = \sum_{i=1}^n \frac{(P_i^R)^2}{n} \quad (9)$$

As shown in Figure 1, the electricity pricing structure is symmetric to the line $\frac{\hat{x}_i}{x_i} = 1$ when $(1 - \frac{0.1}{1.1}) \leq \frac{\hat{x}_i}{x_i} \leq (1 + \frac{0.1}{1.1})$. We can modify the penalty of electricity cost P_i^C in Equation (2) and the penalty of per unit electricity cost P_i^R in Equation (4) to remove this symmetric portion to aggravate the penalty of underestimation, as follows.

$$\text{Modified penalty of electricity cost for month } i, \quad M_i^C = \begin{cases} (\hat{x}_i - x_i) & \text{if } 1 \leq \frac{\hat{x}_i}{x_i} \\ 2(x_i - \hat{x}_i) & \text{if } \frac{\hat{x}_i}{x_i} < 1 \end{cases} \quad (10)$$

$$\text{Modified penalty of per unit electricity cost for month } i, \quad M_i^R = \begin{cases} \frac{\hat{x}_i}{x_i} - 1 & \text{if } 1 \leq \frac{\hat{x}_i}{x_i} \\ 2(1 - \frac{\hat{x}_i}{x_i}) & \text{if } \frac{\hat{x}_i}{x_i} < 1 \end{cases} \quad (11)$$

Notably, $M_i^R = P_i^R$ when $1 \leq \frac{\hat{x}_i}{x_i}$. If $\frac{\hat{x}_i}{x_i} < 1$ (i.e., \hat{x}_i is an underestimation of x_i), then $M_i^R > P_i^R$, indicating a larger penalty using M_i^R than using P_i^R , as shown in the red dash line in Figure 1. Based on M_i^C and M_i^R , another two loss functions are defined as follows.

$$L^{MC} = \sum_{i=1}^n \frac{M_i^C}{n} \quad (12)$$

$$L^{MR} = \sum_{i=1}^n \frac{M_i^R}{n} \quad (13)$$

Equations (10) and (11) consistently incur more penalties to underestimation than to overestimation. In contrast, Equations (6) and (7) put more penalties to underestimation than to overestimation only when the prediction is far from the actual demand, i.e., for underestimation in the range of $(1 - \frac{0.1}{1.1}) > \frac{\hat{x}_i}{x_i}$ and overestimation in the range of $\frac{\hat{x}_i}{x_i} > (1 + \frac{0.1}{1.1})$, to be exact.

5. Performance Study

5.1. Experiment Design

To evaluate the effectiveness of the proposed loss functions, we conducted a performance study using a real dataset from a large manufacturing company in the electronics industry in Taiwan. The dataset contains six time series, corresponding to the company's six power lines. Each time series in the dataset contains the monthly power demands of a power line for 50 consecutive months. Two preprocessing steps were adopted to protect data privacy. First, the monthly power demands in each time series were normalized to between 1 and 2. Second, the 50 consecutive months are numbered from 1 to 50, instead of indicating the exact months and years. The dataset after preprocessing is shown in Figure 2.

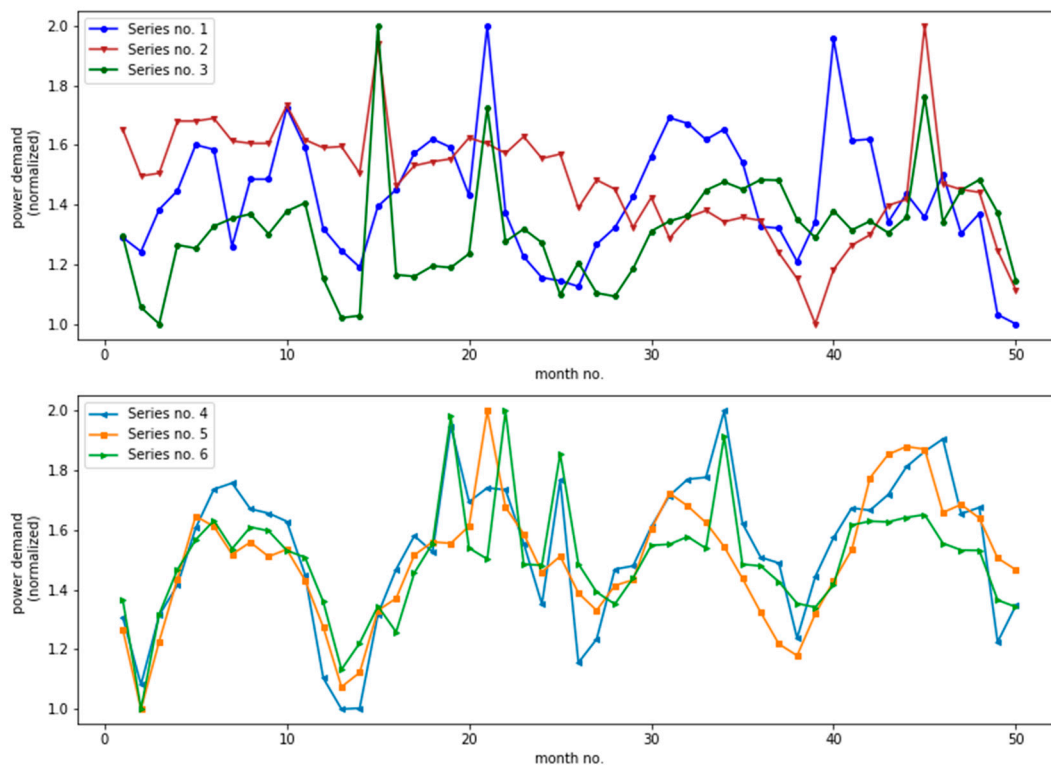


Figure 2. Monthly power demand data (normalized) of six power lines.

Each of the six power lines has its contract capacity. Thus, each time series was handled separately to build its prediction model.

A time-series cross-validation approach was adopted in the experiment for performance evaluation [31]. For each time series $(x_1, x_2, \dots, x_{50})$, a moving window of size 38 was used to yield 13 segments, where the i -th segment contains $(x_i, x_{i+1}, \dots, x_{i+37})$. As illustrated in Figure 3, the window first covers the segment $(x_1, x_2, \dots, x_{38})$, then the segment $(x_2, x_3, \dots, x_{39})$, and so on, and finally the segment $(x_{13}, x_{14}, \dots, x_{50})$. Then, for each segment $(x_i, x_{i+1}, \dots, x_{i+37})$, $i = 1$ to 13, the first 36 elements $(x_i, x_{i+1}, \dots, x_{i+35})$ were used as the training data (shown as blue circles in Figure 3) to build a prediction model, and the last element x_{i+37} was used as the test data (shown as red circles in Figure 3) for performance evaluation. Because the problem under study (see Section 3) requires a

forecast horizon of 2, a model using x_{i+37} as the test data should be not trained with a dataset containing x_{i+36} . Thus, x_{i+36} in the segment $(x_i, x_{i+1}, \dots, x_{i+37})$ was used for neither training nor testing (shown as green circles in Figure 3). Notably, the seasonality of the electronics industry motivates us to use a multiple of 12 months of data for more than two years to train a model. Because each of the time series only contains 50 months of data, we chose to use 36 months of data for training so that we can still retain sufficient data (i.e., $50-36-1=13$ months) for testing. Thus, the size of the moving window is set to 38 (i.e., 36 months for training, one month for gap, and one month for testing).

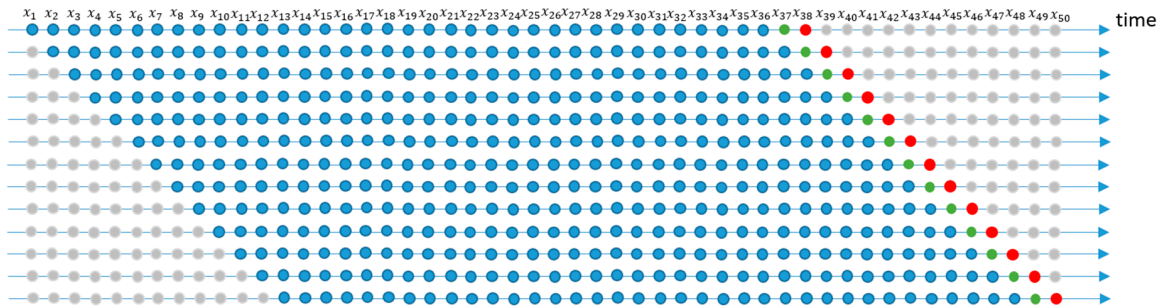


Figure 3. A moving window of size 38 moves on a time series of size 50. For each horizontal time line, the moving window covers from the leftmost blue circle to the red circle.

As shown in Figure 3, a time series in our dataset has 13 segments, and the last element of each segment is used as the test data, so the test data include the last 13 items (i.e., $x_{38}, x_{39}, \dots, x_{50}$) of the time series. Assume that a machine learning scheme predicts the values of $x_{38}, x_{39}, \dots, x_{50}$ as $\hat{x}_{38}, \hat{x}_{39}, \dots, \hat{x}_{50}$. Then, by using the predicted values (i.e., $\hat{x}_{38}, \hat{x}_{39}, \dots, \hat{x}_{50}$) as the contract capacities for their respective months, we can calculate the electricity costs for months 38 to 50 using Equation (1). Then, we evaluate the performance of the machine learning scheme using two performance measures, F_{macro} and F_{micro} , where both are derived from the difference between the electricity cost and the optimal electricity cost, as follows.

$$F_{macro} = \frac{(\sum_{i=38}^{50} C_i) - (\sum_{i=38}^{50} Rx_i)}{\sum_{i=38}^{50} Rx_i} \quad (14)$$

$$F_{micro} = \frac{\sum_{i=38}^{50} \frac{C_i - Rx_i}{Rx_i}}{50 - 38 + 1} \quad (15)$$

Notably, Rx_i is the optimal electricity cost for month i , which occurs when $x_i = \hat{x}_i$. Measure F_{macro} provides a macro view of the performance by using the total electricity cost (i.e., $\sum_{i=38}^{50} C_i$) and the optimal electricity cost (i.e., $\sum_{i=38}^{50} Rx_i$) over the entire period of the test data. In contrast, measure F_{micro} provides a micro view of the performance by first calculating the penalty (i.e., $\frac{C_i - Rx_i}{Rx_i}$) for each month in the period of the test data, and then taking their average. The smaller the values of F_{macro} and F_{micro} , the better the performance. Without loss of generality, we set the basic per unit electricity cost R to 1.

In this experiment, we compared the performance of using LSTM with seven different loss functions: MSE, L^C from Equation (6), L^R from Equation (7), L^{C^2} from Equation (8), L^{R^2} from Equation (9), L^{MC} from Equation (12), and L^{MR} from Equation (13). For each loss function, F_{macro} and F_{micro} were calculated and compared for each time series in the dataset. Then, paired t -test were conducted to check whether using asymmetric functions $L^C, L^R, L^{C^2}, L^{R^2}, L^{MC}$ and L^{MR} is significant better than using the symmetric loss function MSE.

The learning algorithm was implemented using Keras (<https://keras.io/>) in Python. An LSTM neural network model was built with one hidden LSTM layer and one Dense output layer, where the loss function was set to one of the seven loss functions described earlier. A grid search process

was used to determine the hyper-parameters of the model, where three settings (“adam”, “rmsprop” and “nadam”) for the optimizer, three settings (100, 200 and 300) for the learning epochs, and four settings (1, 2, 5 and 10) for the batch learning size were explored. The final month of the learning data was kept aside as the validation data, and the rest of the training data were used to train the model. The validation data were used to evaluate which combination of the hyper-parameters yields the best training performance. Once the best combination of the hyper-parameters was determined, it was adopted to build the LSTM model with all training data. Then the resulting model was used to predict the test data.

5.2. Experimental Results

Tables 1 and 2 shows the values of F_{macro} and F_{micro} for using LSTM with different loss functions, respectively. In most cases, using an asymmetric loss functions (i.e., L^C , L^R , L^{C2} , L^{R2} , L^{MC} or L^{MR}) yields better results (i.e., smaller F_{macro} and F_{micro}) than using the symmetric loss function MSE. There are only six cases where MSE performs better than an asymmetric loss function in terms of F_{macro} (shown in italic in Table 1). Similarly, there are only six cases where MSE performs better than an asymmetric loss function in terms of F_{micro} , as shown in italic in Table 2. Both F_{macro} and F_{micro} show consistent results. Using MSE also yields the worst mean and the second worst standard deviation for both F_{macro} and F_{micro} among the seven loss functions tested. Overall, using an asymmetric loss function results in about 1~2% reduction on F_{macro} and F_{micro} in comparison to using MSE.

Table 1. The values of F_{macro} for using Long Short-Term Memory (LSTM) with different loss functions.

Data	MSE	L^C	L^R	L^{MC}	L^{MR}	L^{C2}	L^{R2}
series 1	23.358%	21.334%	18.504%	21.131%	21.754%	20.815%	20.149%
series 2	28.500%	21.536%	23.041%	21.117%	26.610%	27.598%	28.379%
series 3	8.493%	7.234%	9.446%	7.711%	7.929%	7.839%	4.613%
series 4	11.100%	8.683%	11.121%	11.221%	10.255%	9.318%	10.373%
series 5	10.054%	9.080%	8.476%	7.732%	9.645%	10.825%	8.941%
series 6	7.618%	6.693%	5.748%	6.792%	7.631%	6.813%	7.749%
Mean	14.854%	12.426%	12.723%	12.617%	13.971%	13.868%	13.367%
stdev.	8.815%	7.034%	6.631%	6.761%	8.118%	8.400%	9.035%

Table 2. The values of F_{micro} for using LSTM with different loss functions.

Data	MSE	L^C	L^R	L^{MC}	L^{MR}	L^{C2}	L^{R2}
series 1	22.650%	20.736%	18.191%	20.741%	21.237%	20.946%	19.517%
series 2	26.041%	20.188%	21.200%	19.583%	24.192%	26.031%	27.767%
series 3	8.044%	6.752%	8.730%	7.347%	7.540%	7.354%	4.763%
series 4	11.316%	9.039%	11.932%	11.791%	10.951%	10.157%	11.121%
series 5	10.350%	8.777%	8.194%	8.203%	9.814%	10.478%	8.719%
series 6	7.727%	6.923%	5.930%	7.067%	7.900%	7.053%	8.063%
mean	14.354%	12.069%	12.363%	12.455%	13.606%	13.670%	13.325%
stdev.	7.930%	6.570%	6.070%	6.215%	7.226%	7.898%	8.649%

The results of each asymmetric loss function were compared against the results of MSE using one-tailed paired t -test at significant level 0.05 to check whether the mean of F_{macro} (or F_{micro}) is significantly smaller using an asymmetric loss function than using MSE. The results are shown in Table 3, and only the F_{micro} using L^{MC} or L^{R2} is not significantly smaller than that using MSE.

Table 3. One-tailed Paired *t*-test (significant level $\alpha = 0.05$) against mean squared error (MSE).

		L^C	L^R	L^{MC}	L^{MR}	L^{C2}	L^{R2}
F_{macro}	<i>p</i> -value	0.02461	0.04894	0.048607	0.0157	0.04201	0.04033
	statistic <i>t</i>	−2.5835	−2.03179	−2.0372	−2.9634	−2.152121	−2.1844
	reject H_0	reject H_0	reject H_0	reject H_0	reject H_0	reject H_0	reject H_0
F_{micro}	<i>p</i> -value	0.0137869	0.047644	0.0567112	0.0281908	0.0296349	0.131509
	statistic <i>t</i>	−3.07667	−2.052919	−1.916598	−2.472145	−2.431517	−1.26078
	reject H_0	reject H_0	reject H_0	accept H_0	reject H_0	reject H_0	accept H_0

To examine the prediction performance for each month *i* in the test set, Figure 4 shows the penalty P_i^C on the customer’s electricity cost for month *i* due to under- or overestimation of the actual demand (see Equation (2) for details). The results of using L^{MC} or L^{R2} are excluded from Figure 4 for clarity and because their performance is not significantly different from that of MSE according to the *t*-test in Table 3. Notably, no loss function performs the best for every month in any series.

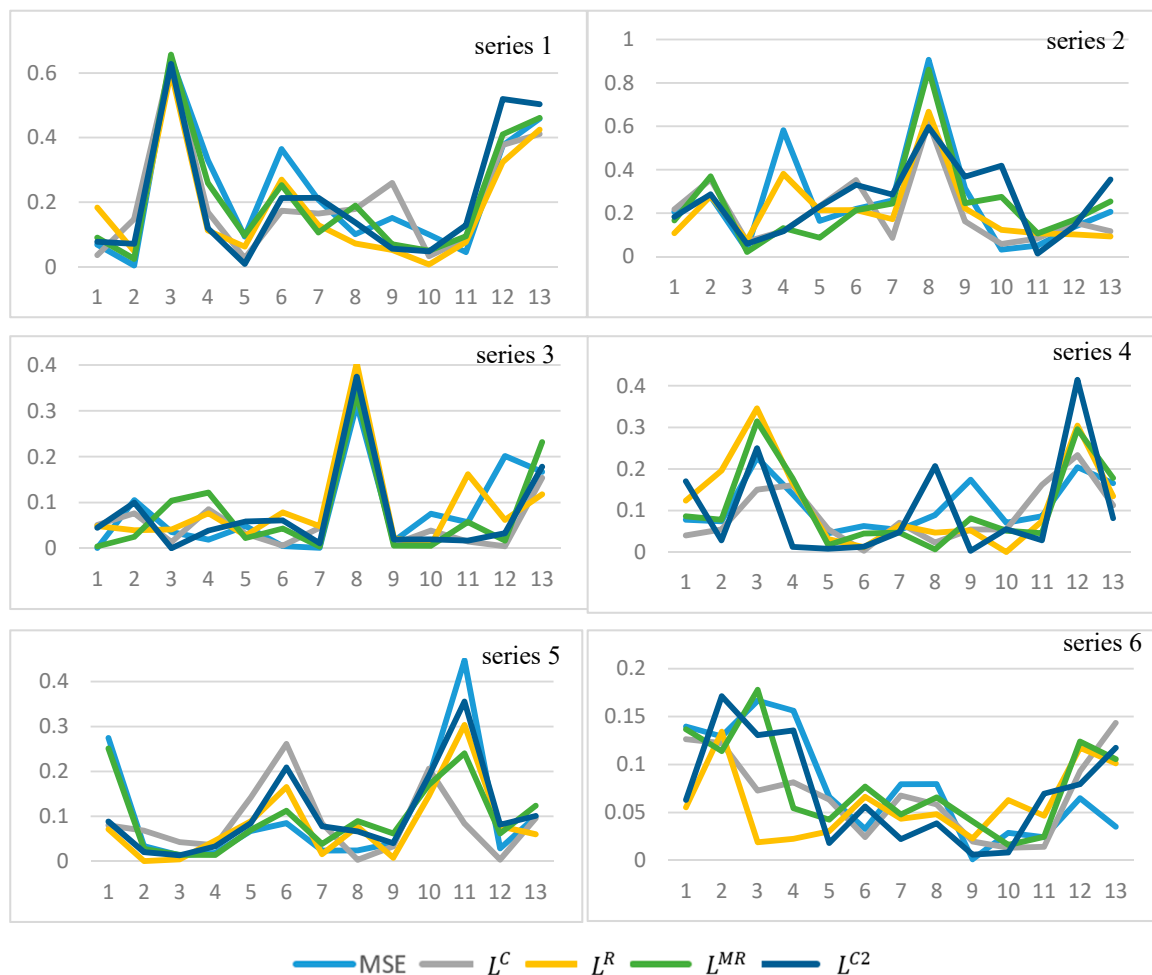


Figure 4. The x-axis is the month no. of the 13 months in the test data, and the y-axis is the penalty P_i^C , which is calculated as the difference between the electricity cost and the optimal cost of month *i* divided by the optimal cost of month *i*.

6. Conclusions

In this study, we built the LSTM models with several asymmetric loss functions derived from the electricity pricing structure. The results show that, in most cases, using the proposed asymmetric

loss functions yields better performance than using the symmetric loss function MSE. Specifically, as shown in Table 3, both F_{macro} and F_{micro} are significantly improved with L^C , L^R , L^{C2} or L^{MR} than with MSE. However, using L^{R2} or L^{MC} only significantly improves F_{macro} but not F_{micro} , and thus L^{R2} and L^{MC} are not good choices for the problem under studied.

Using L^C performs better than using MSE for all six time series, and achieves lower mean F_{macro} value and F_{micro} value than using the other loss functions does (see Tables 1 and 2). Recalled that L^C is based on the electricity cost, while L^R is based on per unit electricity cost. Their simplified versions (i.e., L^{MC} and L^{MR}) and the squared version (i.e., L^{C2} and L^{R2}) do not show much improvement over the original version (i.e., L^C and L^R). Thus, using a loss function that directly reflects the electricity cost is a good choice to start with.

The main contribution of this study is two-fold. First, we showed that using the LSTM model with a loss function consistent with the electricity pricing structure can reduce the electricity cost. Second, although we only experimented with the LSTM model in this study, the same idea can be easily adapted to other machine learning algorithms that use a loss function to adjust the prediction model during the learning process.

The electricity pricing structure may be different in different countries and regions. It remains unknown whether our results can be applied directly to the other electricity pricing structures. However, most pricing structures are asymmetric with respect to under- and overestimation of the power demand. Although some pricing structures may add more (or less) penalty to underestimating power demands, adapting the loss function to the electricity pricing structure is a direction with great potential for reducing the electricity cost.

Several directions are worth pursuing in future research. First, we only experimented with the LSTM to build the prediction models. Future research can explore other machine learning techniques to improve prediction performance further. Second, developing a more sophisticated loss function that takes into account both the problem goal (i.e., lower electricity cost) and the convergence performance of the machine learning techniques can be further explored.

Author Contributions: Conceptualization, Methodology, Data curation, Writing, Supervision, Visualization, Funding acquisition, J.-L.L.; Validation, Software, Y.Z., K.Z., B.C. and F.Z. Overall contribution: J.-L.L. (60%), Y.Z. (10%), K.Z. (10%), B.C. (10%) and F.Z. (10%). All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Ministry of Science and Technology (MOST), Taiwan, under Grant MOST 108-2221-E-155-013.

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

Nomenclature

R	Basic per unit electricity cost
\hat{x}_i, x_i	Contract capacity and actual power demand for month i , respectively
C_i, R_i	Electricity cost and per unit electricity cost for month i , respectively
P_i^C	Penalty on the electricity cost for month i
P_i^R	Penalty on the per unit electricity cost for month i
M_i^C	Modified penalty of electricity cost for month i
M_i^R	Modified penalty of per unit electricity cost for month i
$L^C, L^R, L^{C2}, L^{R2}, L^{MC}, L^{MR}$	Loss functions based on $P_i^C, P_i^R, (P_i^C)^2, (P_i^R)^2, M_i^C$ and M_i^R , respectively.
F_{macro}	Percentage of deviation from the optimal total electricity cost
F_{micro}	Average of the percentage of deviation from the optimal monthly electricity cost

References

1. Papalexopoulos, A.D.; Hesterberg, T.C. A regression-based approach to short-term system load forecasting. *IEEE Trans. Power Syst.* **1990**, *5*, 1535–1547. [[CrossRef](#)]

2. Xia, C.; Wang, J.; McMenemy, K. Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. *Int. J. Electr. Power Energy Syst.* **2010**, *32*, 743–750. [[CrossRef](#)]
3. Amjady, N.; Keynia, F. Mid-term load forecasting of power systems by a new prediction method. *Energy Convers. Manag.* **2008**, *49*, 2678–2687. [[CrossRef](#)]
4. Gross, G.; Galiana, F.D. Short-term load forecasting. *Proc. IEEE* **1987**, *75*, 1558–1573. [[CrossRef](#)]
5. Srivastava, A.K.; Pandey, A.S.; Singh, D. Short-term load forecasting methods: A review. In Proceedings of the 2016 International Conference on Emerging Trends in Electrical Electronics & Sustainable Energy Systems (ICETEESES), Sultanpur, India, 11–12 March 2016; pp. 130–138.
6. Santra, A.S.; Lin, J.-L. Integrating Long Short-Term Memory and Genetic Algorithm for Short-Term Load Forecasting. *Energies* **2019**, *12*, 2040. [[CrossRef](#)]
7. Coakley, D.; Raftery, P.; Keane, M. A review of methods to match building energy simulation models to measured data. *Renew. Sustain. Energy Rev.* **2014**, *37*, 123–141. [[CrossRef](#)]
8. Bianco, V.; De Rosa, M.; Scarpa, F.; Tagliafico, L.A. Analysis of energy demand in residential buildings for different climates by means of dynamic simulation. *Int. J. Ambient Energy* **2016**, *37*, 108–120. [[CrossRef](#)]
9. Pallonetto, F.; Mangina, E.; Milano, F.; Finn, D.P. SimApi, a smartgrid co-simulation software platform for benchmarking building control algorithms. *SoftwareX* **2019**, *9*, 271–281. [[CrossRef](#)]
10. Rout, U.K.; Voß, A.; Singh, A.; Fahl, U.; Blesl, M.; Gallachóir, B.P.Ó. Energy and emissions forecast of China over a long-time horizon. *Energy* **2011**, *36*, 1–11. [[CrossRef](#)]
11. Sánchez-Durán, R.; Luque, J.; Barbancho, J. Long-Term Demand Forecasting in a Scenario of Energy Transition. *Energies* **2019**, *12*, 3095. [[CrossRef](#)]
12. Baharlouei, Z.; Hashemi, M.; Narimani, H.; Mohsenian-Rad, H. Achieving Optimality and Fairness in Autonomous Demand Response: Benchmarks and Billing Mechanisms. *IEEE Trans. Smart Grid* **2013**, *4*, 968–975. [[CrossRef](#)]
13. Jacob, M.; Neves, C.; Vukadinović Greetham, D. Short Term Load Forecasting. In *Forecasting and Assessing Risk of Individual Electricity Peaks*; Springer International Publishing: Cham, Switzerland, 2020; pp. 15–37. [[CrossRef](#)]
14. Zheng, H.; Yuan, J.; Chen, L. Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation. *Energies* **2017**, *10*, 1168. [[CrossRef](#)]
15. Wang, Y.; Velswamy, K.; Huang, B. A Long-Short Term Memory Recurrent Neural Network Based Reinforcement Learning Controller for Office Heating Ventilation and Air Conditioning Systems. *Processes* **2017**, *5*, 46. [[CrossRef](#)]
16. Abu-Shikhah, N.; Elkarmi, F. Medium-term electric load forecasting using singular value decomposition. *Energy* **2011**, *36*, 4259–4271. [[CrossRef](#)]
17. Bozkurt, Ö.Ö.; Biricik, G.; Taysi, Z.C. Artificial neural network and SARIMA based models for power load forecasting in Turkish electricity market. *PLoS ONE* **2017**, *12*, e0175915. [[CrossRef](#)]
18. Liu, Y.; Wang, W.; Ghadimi, N. Electricity load forecasting by an improved forecast engine for building level consumers. *Energy* **2017**, *139*, 18–30. [[CrossRef](#)]
19. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. *arXiv* **2014**, arXiv:1409.3215.
20. Ordóñez, J.F.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115. [[CrossRef](#)]
21. Liu, Q.; Zhou, F.; Hang, R.; Yuan, X. Bidirectional-Convolutional LSTM Based Spectral-Spatial Feature Learning for Hyperspectral Image Classification. *Remote Sens.* **2017**, *9*, 1330. [[CrossRef](#)]
22. Hashemi, S.R.; Salehi, S.S.M.; Erdogmus, D.; Prabhu, S.P.; Warfield, S.K.; Gholipour, A. Asymmetric Loss Functions and Deep Densely-Connected Networks for Highly-Imbalanced Medical Image Segmentation: Application to Multiple Sclerosis Lesion Detection. *IEEE Access* **2019**, *7*, 1721–1735. [[CrossRef](#)] [[PubMed](#)]
23. Akkad, K.; He, D. A Hybrid Deep Learning Based Approach for Remaining Useful Life Estimation. In Proceedings of the 2019 IEEE International Conference on Prognostics and Health Management (ICPHM), San Francisco, CA, USA, 17–20 June 2019; pp. 1–6.
24. Ambach, D.; Vetter, P. Wind Speed and Power Forecasting—A Review and Incorporating Asymmetric Loss. In Proceedings of the 2016 Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO), Beer-Sheva, Israel, 15–18 February 2016; pp. 115–123.

25. Chen, H.; Wan, Q.; Wang, Y. Refined Diebold-Mariano Test Methods for the Evaluation of Wind Power Forecasting Models. *Energies* **2014**, *7*, 4185–4198. [[CrossRef](#)]
26. Pierdzioch, C.; Rülke, J.-C. A Note on Forecasting the Rate of Change of the Price of Oil: Asymmetric Loss and Forecast Rationality. *Economies* **2013**, *1*, 6–13. [[CrossRef](#)]
27. Chen, C.-Y.; Liao, C.-J. A linear programming approach to the electricity contract capacity problem. *Appl. Math. Model.* **2011**, *35*, 4077–4082. [[CrossRef](#)]
28. Tsay, M.T.; Lin, W.M.; Lee, J.L. Optimal contracts decision of industrial customers. *Int. J. Electr. Power Energy Syst.* **2001**, *23*, 795–803. [[CrossRef](#)]
29. Lee, T.-Y.; Chen, C.-L. Iteration particle swarm optimization for contract capacities selection of time-of-use rates industrial customers. *Energy Convers. Manag.* **2007**, *48*, 1120–1131. [[CrossRef](#)]
30. Ferdavani, A.K.; Gooi, H.B. The very fast method for contracted capacity optimization problem in Singapore. In Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Singapore, 22–25 November 2016; pp. 2100–2103.
31. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Melbourne, Australia, 2018.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).