DEEP REINFORCEMENT LEARNING BASED RENEWABLE ENERGY ERROR COMPENSABLE FORECASTING

Jaeik Jeong & Hongseok Kim
Department of Electronic Engineering, Sogang University, Seoul 04107, South Korea
{jaeik1213,hongseok}@sogang.ac.kr

ABSTRACT

Recently, renewable energy is rapidly integrated into the power grid to prevent climate change, and accurate forecasting of renewable generation becomes critical for reliable power system operation. However, existing forecasting algorithms only focused on reducing forecasting errors without considering error compensability by using a large-scale battery. In this paper, we propose a novel strategy called *error compensable forecasting*. We switch the objective of forecasting from reducing errors to making errors compensable by leveraging a battery. Specifically, we propose a deep reinforcement learning based framework having forecasting in the loop of control. Extensive simulations show that the proposed one-hour ahead forecasting achieves zero error for more than 98% of time while reducing the operational expenditure by up to 44%.

1 INTRODUCTION

The Paris Agreement has recently stressed the necessity of using renewable energy instead of fossil fuels to prevent climate change. As a result, global penetration of renewable energy rapidly increases, but renewable power outputs heavily depend on weather conditions such as clouds, temperature, and humidity. This causes substantial uncertainties, which brings adverse effects on economic benefit and the stability of power grids. Thus, accurate renewable energy forecasting techniques are required to integrate the renewable energy into the power grids and ultimately prevent climate change. Recently, deep learning based renewable generation forecasting techniques have been proposed (Cardona et al., 2019; Mathe et al., 2019; Jeong & Kim, 2019) and show significantly improved performances compared to conventional machine learning based schemes. However, forecasting always induces errors, and large-scale energy storages such as lithium-ion batteries are used to compensate forecasting errors (Bae et al., 2016; Gholami et al., 2018; Bae et al., 2019). The basic idea is such that over-forecasting errors are compensated by discharging energy from the battery while under-forecasting errors are resolved by charging excessive generation into the battery.

Traditional forecasting methods commonly aimed to minimize the forecasting errors. They used the mean squared error (MSE) as an objective function in training process. Since there is a squared term in the MSE, training is processed without considering whether the errors are positive or negative. However, reducing errors does not necessarily imply compensable errors. For example, suppose that the battery is empty, i.e., discharging is not possible but charging battery is possible. Then, over-forecasting is not allowed because the battery cannot compensate error by discharging. By contrast, under-forecasting is compensable because excessive generation can be stored in the battery. For the same reason, under-forecasting is not allowed when the battery is fully charged. Nevertheless, existing forecasting algorithms do not consider whether the errors are positive or negative but just reduce the distance between forecasting and real values. Consequently, none of the previous works considered error compensability by using the battery.

In this regard, we propose a novel strategy called *error compensable forecasting* (ECF). We switch the objective of forecasting from reducing errors to making compensable errors. The challenging part of developing ECF lies in that the stored energy at current time is affected by the previous forecasting result. Hence, time-coupling exists between forecasting and battery control, and forecasting should be in the loop of *sequential decision making*. We tackle this problem by leveraging reinforcement learning (RL) that has interaction between an agent and the environment where actions of the agent affect the subsequent data it receives (Sutton et al., 1998). In our framework, an action is a continuous forecasted value, and it requires a continuous action space. To enable the continuous control, we leverage the state-of-the-art deep reinforcement learning (DRL) algorithm called proximal policy optimization (PPO), which is known to be simpler to implement than other DRL algorithms with outstanding performance (Schulman et al., 2017). Our extensive simulations with real solar and wind power generation data confirm that the proposed framework outperforms the traditional forecasting and achieves zero error for more than 98% of time for one-hour ahead forecasting when the maximum battery capacity is 0.5 p.u, i.e., a half of the installed generation capacity.

2 METHODS

2.1 BATTERY OPERATION

In this section, we present a practical battery model for ECF applications. For simplicity we focus on one-hour ahead forecasting because renewable energy providers can resubmit their bids one-hour ahead of the operation hour in a renewable energy market (Bae et al., 2019). Since battery degradation is known to be severe at both ends of the state-of-charge (SoC), i.e., either empty or full, the stored energy denoted by E_t at time slot t should be constrained by (Choi & Kim, 2016)

$$E_{\max} \cdot \mathrm{SoC}_{\min} \leq E_t \leq E_{\max} \cdot \mathrm{SoC}_{\max}, \tag{1}$$

where $\mathrm{SoC}_{\min}$ and $\mathrm{SoC}_{\max}$ denote the minimum and maximum SoC of the battery, and $E_{\max}$ denotes the maximum battery capacity. From (1), the charging and discharging power limitation at time slot t, denoted by $\bar{P}_t^c$ and $\bar{P}_t^d$, can be obtained as

$$\bar{P}_t^c = \min\left(P_{\max}^c, \quad \frac{1}{\eta_c} \cdot \frac{E_{\max} \cdot \mathrm{SoC}_{\max} - E_t}{\Delta t}\right), \tag{2a}$$

$$\bar{P}_t^d = \min\left(P_{\max}^d, \quad \eta_d \cdot \frac{E_t - E_{\max} \cdot \mathrm{SoC}_{\min}}{\Delta t}\right), \tag{2b}$$

where Δt is the duration of time slot, and η_c, η_d are the charging and discharging efficiencies, respectively, and $P_{\max}^c$, $P_{\max}^d$ are the maximum charging power and discharging power of the battery, respectively, which are inherently determined by the power conditioning system constraints.

Let x_t be the real generation value in time slot t, and a_t be the *forecasted value* in the next time slot $t+1$. When a_t is smaller than x_{t+1} (under-forecasting), excessive energy $x_{t+1} - a_t$ is stored in the battery to match the forecasted value, but it is limited to $\bar{P}_t^c$ as in (2a). Likewise, when a_t is higher than x_{t+1} (over-forecasting), energy deficit $a_t - x_{t+1}$ is compensated by drawing energy from the battery, up to $\bar{P}_t^d$ as in (2b). Accordingly, in the next time slot $t+1$, the charging and discharging power of the battery, denoted by P_{t+1}^c and P_{t+1}^d, are formulated as

$$P_{t+1}^c = \min\left((x_{t+1} - a_t)^+, \bar{P}_t^c\right), \tag{3a}$$

$$P_{t+1}^d = \min\left((a_t - x_{t+1})^+, \bar{P}_t^d\right), \tag{3b}$$

where $(x)^+ = \max(x, 0)$. Accordingly, E_t evolves in time as follows:

$$E_{t+1} = E_t + \eta_c P_{t+1}^c \Delta t - \frac{1}{\eta_d} P_{t+1}^d \Delta t. \tag{4}$$

2.2 ERROR COMPENSATION COST

We mainly consider three costs from battery degradation, energy transfer loss, and non-compensable errors. First, degradation cost roughly proportional to the charging and discharging power if (1) is satisfied (Han et al., 2014; Kim et al., 2016). Second, charging efficiency (η_c) and discharging efficiency (η_d) are not perfect, which causes energy transfer loss. Thus, the battery cost at the next time slot $t+1$, denoted by B_{t+1}, is the sum of these two costs:

$$B_{t+1} = b\left(P_{t+1}^c + P_{t+1}^d\right)\Delta t + l\left[\left(1 - \eta_c\right)P_{t+1}^c + \left(\frac{1}{\eta_d} - 1\right)P_{t+1}^d\right]\Delta t, \tag{5}$$

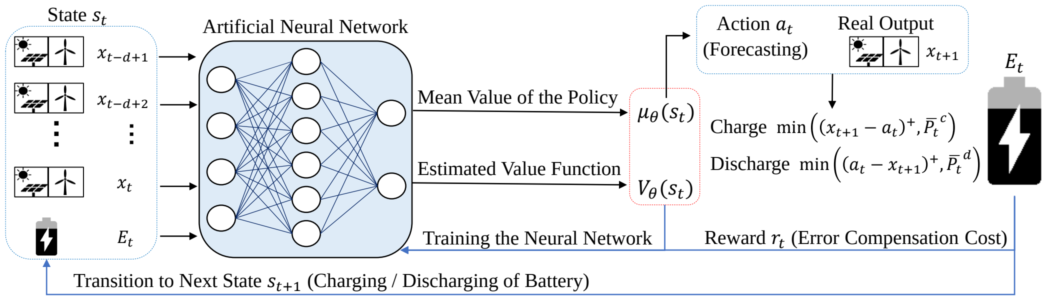

Figure 1: A framework of DRL-based error compensable forecasting.

where b is the degradation cost per unit energy, and l is the penalty for energy loss per unit energy.

Next, we consider the cost from non-conpensable errors. When excessive energy cannot be stored in the battery, the power system operator need to curtail the output power, which causes energy loss, and when energy deficit cannot be compensated by the battery, the power system operator purchases power from a reserve market (Kim & Powell, 2011; Ryu et al., 2018). Thus, the cost from non-compensable errors at the next time slot $t+1$, denoted by N_{t+1}, is given by:

$$N_{t+1} = l\left(x_{t+1} - a_t - P^c_{t+1}\right)^+ \Delta t + p\left(a_t - x_{t+1} - P^d_{t+1}\right)^+ \Delta t, \tag{6}$$

where p is the power purchasing cost per unit energy.

We then formulate our problem as optimization as follows.

$$\begin{aligned} \text{minimize} \quad & \lim_{T\to\infty} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t \left(B_{t+1} + N_{t+1}\right)\right], \\ \text{subject to} \quad & (2a), (2b), (3a), (3b), \text{and } (4), \\ \text{variables} \quad & \{a_t\}_{t=0}^{\infty}, \end{aligned} \tag{7}$$

where $\gamma \in (0,1)$ is a discounted factor that determines the importance of future costs. It is obvious that the optimal solution is $a_t = x_{t+1}, \forall t$, when error compensation is not needed. However, x_{t+1} is unknown at time slot t, and forecasting is required to solve the problem (7). In general, η_c and η_d are generally close to 1, and degradation cost is much less than the profit loss and the power purchasing cost, which implies that b is much less than l and p (Kim et al., 2019), see Table 1.

2.3 Deep Reinforcement Learning based Solution

In problem (7), time-coupling exists because of the equation (4). Hence, the forecasting in our problem is essentially sequential decision making under uncertainty. In this regard, we consider RL with a set of states $\mathcal{S}$ and a set of actions $\mathcal{A}$. At time slot t, an agent takes an action $a_t \in \mathcal{A}$ at state $s_t \in \mathcal{S}$ and goes to a next state $s_{t+1} \in \mathcal{S}$ with a reward r_{t+1}. The solution is determined by x_{t+1} and E_t, but as x_{t+1} is unknown at time t, we use the observed past d values as in the time-series forecasting. As a result, we set the state $s_t = (x_{t-d+1}, x_{t-d+2}, \cdots, x_t, E_t)$, the action a_t as the forecasted value in the next time slot $t+1$, and the reward $r_t = -(B_t + N_t)$.

DRL combines the classic RL with the deep neural network (DNN), which is also suitable for problems considering continuous state and action space, which is the main interest of this paper. The policy $\pi_\theta(a_t|s_t)$ is generally captured by Gaussian distribution with the parameters θ. One network, called actor, outputs its mean $\mu_\theta(s_t)$ with variable standard deviations. Or, one can use a predetermined small value of standard deviation σ to improve stability (Zimmer & Weng, 2019). The other network, called critic, outputs the *estimated* value function $V_\theta(s_t)$ to estimate the value function accurately. In practice, all parameters of non-output layers can be shared in actor and critic, so we use one DNN to generate $\mu_\theta(s_t)$ and $V_\theta(s_t)$. To train the DNN, we apply PPO, which is known to be much simpler to implement than other DRL algorithms with outstanding performance, where the details of the training process are elaborated in the (Schulman et al., 2017). The overall architecture of the proposed DRL based ECF is shown in Figure 1.

Table 1: Battery related parameters

Parameters	Δt	η_c	η_d	$\mathrm{SoC}_{\min}$	$\mathrm{SoC}_{\max}$	$P^c_{\max}/E_{\max}$	$P^d_{\max}/E_{\max}$	b	l	p	γ
Value	1 hour	0.9	0.9	0.1	0.9	1/3	1/3	\$10/MWh	\$50/MWh	\$100/MWh	0.99

Table 2: Experiment results (solar)

	$E_{\max} = 0.25$ p.u.		$E_{\max} = 0.5$ p.u.	
	BF	ECF	BF	ECF
MAPE	18.74%	10.08%	17.70%	0.13%
Score	0.729	0.848	0.765	0.990
Mean Cost	\$2593	\$2260	\$2430	\$1455

Table 3: Experiment results (wind)

	$E_{\max} = 0.25$ p.u.		$E_{\max} = 0.5$ p.u.	
	BF	ECF	BF	ECF
MAPE	6.16%	1.21%	4.85%	0.20%
Score	0.642	0.883	0.734	0.983
Mean Cost	\$2737	\$1803	\$2368	\$1332

3 RESULTS

In this section, we evaluate the performance of the proposed ECF. We compare our models with the baseline forecasting (BF) that determines a_t by training DNN with the MSE between x_{t+1}. We use two real-world open datasets, aggregated production of solar power and wind power across Belgium from January 1st 2016 to December 31th 2019, released by Elia[1]. We normalize the data between 0 and 1 by the installed renewable generation capacity (3887MW for solar and 3796MW for wind) and sample every 1 hour. We split the dataset into training set (50%, two years), validation set (25%, one year), and test set (25%, one year) in chronological order. In solar power datasets, we exclude the data during night (zero-value data). For the validation and testing phases in ECF, we built a deterministic PPO defined in (Zimmer & Weng, 2019). We also normalize the capacity of the battery by the installed renewable generation capacity to obtain per unit (p.u.) quantity. At time slot $t = 0$, we set $E_t = 0.5 \times E_{\max}$, i.e., the half stored energy, and E_t for $t \geq 1$ is determined by (4). The battery related parameters are from (Kim & Powell, 2011; Ryu et al., 2018; Kim et al., 2019) and summarized in Table 1.

We use a fully-connected multi-layer perceptron based architecture as it is one of the attractive solutions for one-hour ahead forecasting problems (Bae et al., 2016; Gholami et al., 2018; Bae et al., 2019) with a shorter training time than recurrent neural network (Goodfellow et al., 2016). We select 4 input neurons in BF case (where $d = 4$) and 5 in ECF case (to include E_t), two hidden layers and 16 neurons per each layer based on the validation set.

We use the mean absolute percentage error (MAPE) to evaluate the performance of each technique. Since the compensated real output is determined by charging and discharging the battery, the MAPE in our case is defined as

$$\mathrm{MAPE} = \frac{100}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \left| \frac{\left(x_{t+1} - P^c_{t+1} + P^d_{t+1}\right) - a_t}{\left(x_{t+1} - P^c_{t+1} + P^d_{t+1}\right)} \right| [\%], \tag{8}$$

where $\mathcal{T}$ is a test dataset. Also, to evaluate error compensability, we evaluate the score, the ratio of the time slots where the errors are completely compensated by battery, and the mean cost, the mean value of $B_t + N_t$. Table 2 and Table 3 show the performances of the BF and ECF when $E_{\max}$ is 0.25 p.u. and 0.5 p.u. for the solar and wind datasets, respectively. The proposed ECF far improves all the performances compared to the BF. Furthermore, when $E_{\max} = 0.5$ p.u., the improvements become significant, e.g., the MAPE becomes near zero, and the score becomes 0.99 (solar) and 0.983 (wind), which implies that ECF achieves zero error for more than 98% of time.

4 CONCLUSION

In this paper, we proposed a novel forecasting strategy called ECF for renewable energy where the objective is switched from reducing errors to making compensable errors by using battery. The proposed model shows significantly better performance than the traditional forecasting in the sense of error compensability. Future research can be extended into multi-step ahead (such as day ahead) forecasting algorithm with evaluating the economical impact such as day ahead bidding profit.

[1]http://www.elia.be/en/grid-data/power-generation

Acknowledgments

This work was supported in part by Smart City R&D project of the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport under Grant 19NSPS-B152996-02.

References

Kuk Yeol Bae, Han Seung Jang, and Dan Keun Sung. Hourly solar irradiance prediction based on support vector machine and its error analysis. *IEEE Transactions on Power Systems*, 32(2): 935–945, 2016.

Kuk Yeol Bae, Han Seung Jang, Bang Chul Jung, and Dan Keun Sung. Effect of prediction error of machine learning schemes on photovoltaic power trading based on energy storage systems. *Energies*, 12(7):1249, 2019.

Jennifer Cardona, Michael Howland, and John Dabiri. Seeing the wind: Visual wind speed prediction with a coupled convolutional and recurrent neural network. In *Advances in Neural Information Processing Systems*, pp. 8732–8742, 2019.

Yohwan Choi and Hongseok Kim. Optimal scheduling of energy storage system for self-sustainable base station operation considering battery wear-out cost. *Energies*, 9(6):462, 2016.

Mehrdad Gholami, Seyed Hamid Fathi, Jafar Milimonfared, Zhe Chen, and Fujin Deng. A new strategy based on hybrid battery–wind power system for wind power dispatching. *IET Generation, Transmission & Distribution*, 12(1):160–169, 2018.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

Sekyung Han, Soohee Han, and Hirohisa Aki. A practical battery wear model for electric vehicle charging applications. *Applied Energy*, 113:1100–1108, 2014.

Jaeik Jeong and Hongseok Kim. Multi-site photovoltaic forecasting exploiting space-time convolutional neural network. *Energies*, 12(23):4490, 2019.

Hongseok Kim, Joohee Lee, Shahab Bahrami, and Vincent Wong. Direct energy trading of microgrids in distribution energy market. *IEEE Transactions on Power Systems*, 2019.

Jae Ho Kim and Warren B Powell. Optimal energy commitments with storage and intermittent supply. *Operations research*, 59(6):1347–1360, 2011.

Kangsan Kim, Yohwan Choi, and Hongseok Kim. Data-driven battery degradation model leveraging average degradation function fitting. *Electronics Letters*, 53(2):102–104, 2016.

Johan Mathe, Nina Miolane, Nicolas Sebastien, and Jeremie Lequeux. PVNet: A LRCN architecture for spatio-temporal photovoltaic power forecasting from numerical weather prediction. *arXiv preprint arXiv:1902.01453*, 2019.

Seunghyoung Ryu, Sunghwan Bae, Jong-Uk Lee, and Hongseok Kim. Gaussian residual bidding based coalition for two-settlement renewable energy market. *IEEE Access*, 6:43029–43038, 2018.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Richard S Sutton, Andrew G Barto, et al. *Introduction to reinforcement learning*. MIT press Cambridge, 1998.

Matthieu Zimmer and Paul Weng. Exploiting the sign of the advantage function to learn deterministic policies in continuous domains. *arXiv preprint arXiv:1906.04556*, 2019.