Proceedings of the 2020 IISE Annual Conference
L. Cromarty, R. Shirwaiker, P. Wang, eds.

Data-Driven Optimal Control of Wind Turbines Using Model-Based Reinforcement Learning

Abstract ID: 790175

Shenglin Peng, Qianmei Feng
Department of Industrial Engineering, University of Houston
Houston, TX 77204, USA

Abstract

We propose a model-based reinforcement learning approach for maximizing the power output of wind turbines (WTs). The optimal control of wind turbines majorly uses the maximum power point tracking (MPPT) strategy for sequential decision-making that can be modeled as a Markov decision process (MDP). In the literature, the continuous control variables are typically discretized to cope with the curse of dimensionality in traditional dynamic programming methods. To provide more accurate prediction, we formulate the problem into a continuous state space, continuous action space MDP by utilizing the function approximation in reinforcement learning. The commonly used pitch angle and torque are considered as control variables, which are regarded as the system state along with some other uncontrollable variables proven to affect the power output. Computational studies of real data are used to demonstrate that the proposed method outperforms the existing methods in the literature in obtaining the optimal power output.

Keywords
Markov decision process; reinforcement learning; function approximation; wind turbines

1. Introduction

Wind energy is considered as a promising alternative energy source because it is renewable, cost-effective, and environmental friendly [1]. The capability of wind power as a competitive energy source is evidenced by its drastic growth. From 2001 to 2018, the cumulative capacity of installed wind power worldwide has grown from 24GW to 591GW. More than 50GW wind capacity has been installed annually since 2014, according to the Global Wind Energy Council [2]. Along with the rapid growth of wind energy capacity, there is also a growth in the size and power output of a single wind turbine. The utility-scale wind turbines, starting from a height of 24 meters and 50kW of output, have become as large as 114-meter high and 5GW of power output [3], which has been motivated by the economic advantages of large wind turbines. The growing size of wind turbines makes this industry highly capital-sensitive, in which a small fraction of the decrease in power output and operation time could lead to significant monetary loss. With the average price of electricity assumed to be around $0.1 per kWh [4], even 1% energy loss on a 100MW wind turbine is estimated to reduce the annual revenue by $307,500 [3].

In such a capital-intensive industry, owners of large wind turbines can benefit greatly by optimizing operation and maintenance of wind turbines. Actually, megawatt-scale wind turbines with variable speed become particularly attractive due to their cost-effectiveness stemming from their operation that can be actively controlled [5]. The optimal control of wind turbines majorly uses the maximum power point tracking (MPPT) to directly increase the power output when the wind profile deviates from the standard point. To extend the total service life of wind turbines, the structural load reduction is implemented to help reduce the productivity loss caused by the maintenance.

Different modeling and optimization algorithms have been used in MPPT with various control variables. The estimation of power output adopts the linear regression with polynomial features and time-series models such as Kalman filter [6]. With the deployment of supervisory control and data acquisition (SCADA) system [7], modern data-driving techniques such as artificial neural networks (ANN) have also been utilized [8]. The generator torque is the main control variable in MPPT research as it is directly related to the power generated, while the pitch angle is also adjusted to capture the maximal amount of wind power under a changing wind profile [9]. In most optimization models, the output power is maximized at a single time point, which is less practical due to the time lag between the

observation of signals and the optimization decision-making. Moreover, such algorithms fail to take into account the correlation between consecutive control decisions, which makes it difficult to incorporate the constraints of the maximum changing rate of control variables.

As a sequential decision-making task, MPPT of wind turbines can naturally fit into the model of Markov decision processes (MDPs). For discrete and small state and action spaces, the linear programming and dynamic programming (DP) solutions to such MDP problems were developed in the early years [10]. The DP methods use a table to represent all state values (or state-action values) and are thus called tabular methods. The DP methods suffer a lot from the curse of dimensionality that makes them unable to handle large, high-dimensional or continuous state and action sets. Recently, researchers approached the MPPT problem as an MDP by discretizing a control variable (i.e., turbine rotating speed) in order to fit in the tabular Q learning, which inevitably introduces the discretization error [11].

In this research, we aim to maximize the WT power output under the stochastic wind profile by formulating the problem as a continuous state space, continuous action space Markov decision process, without discretizing the variables. Instead, the curse of dimensionality of DP methods is overcome by utilizing the function approximation in reinforcement learning. Reinforcement learning is a modern MDP solving process that approaches large MDPs when exact methods become infeasible. As an important technique in reinforcement learning, function approximation uses a function to approximate state values or state-action values, and bootstrap from previous approximated value functions to carry out DP iterations. In this research, we explore different function approximations on the Q-function, with an aim to find the optimal control rule with undiscounted reward and infinite horizon. A fitted Q iteration algorithm is used in the framework of off-policy reinforcement learning.

2. Physical Mechanisms of Wind Turbines

Wind turbines generate the electrical power by capturing the wind power. The wind power that can be extracted by a wind turbine is generally given by [3, 8, 13]

$$P = P_{wind} C_p(\lambda, \beta) = \frac{1}{2} \rho \pi R^2 v^3 C_p(\lambda, \beta) \tag{1}$$

where P_{wind} is the theoretical wind power available to a turbine, ρ is the air density, R is the rotor radius, and v is the wind speed before passing the rotor. $C_p(\lambda, \beta)$ is the power coefficient that evaluates the proportion of available wind power captured by the wind turbine, which is a nonlinear function of blade pitch angle β and the tip-speed ratio $\lambda = \omega_r R/v$ where ω_r the rotational speed of the rotor [8].

The function C_p is turbine specific that can be estimated by field experiments and/or specialized simulation. It usually has a maximum value at the optimal blade pitch angle β^* and tip-speed ratio λ^*, which are provided for each specific turbine. Therefore, the power from a wind turbine can be controlled by the rotor speed ω_r, in addition to the blade pitch angle β.

3. Methodology

The optimal control problems of wind turbines can be approached by using reinforcement learning techniques, where the problem is formulated as a Markov decision process.

3.1. MDP and Reinforcement Learning

A stationary MDP is characterized by a quintuple $\{T, S, A_s, p(\cdot|s, a), r(s, a)\}$ consisting of the set of decision epochs T, a state space S, an action space A_s under state s, a stochastic transition function p, and a reward function r [14]. In each decision epoch, an action available for the current state s is selected and an instant reward r(s, a) is received. The probability distribution of the next state $p(\cdot|s, a)$ completely depends on the current state and action, which is a core assumption of MDPs.

In an MDP model, a controller or agent seeks to find a policy $\pi: S \rightarrow A$ that maximizes a certain value criterion related to the reward. A commonly used criterion is the expected total discounted reward

$$J_\infty^\pi(s) = \lim_{h \to \infty} \sum_{t=0}^{h} E[\gamma^t r(s_t, \pi(s_t)) | s_0 = s], \tag{2}$$

where γ is the discount factor. When the reward represents a revenue, it is proper to choose γ as the reciprocal of the risk-free rate. Such a discounted reward is referred to as the value function, or the V-function, of the state s under the policy π, denoted by $v_\pi(s)$. The goal of MDP is to find an optimal policy π^* such that

$$v_{\pi*}(s) \geq v_\pi(s) \ \ \forall s \in S, \pi. \tag{3}$$

In most cases, we write $v_{\pi^*}(s)$ as $v_*(s)$ for brevity. The value function under the optimal policy should satisfy the Bellman optimality equation, a central property of MDPs [11]:

$$v_*(s) = \max_a \sum_{s',r} p(s', r|s, a)[r + \gamma v_*(s')]. \tag{4}$$

In general, the optimization problem using the V-function is computationally intractable, unless an explicit model is assumed. The machine learning approach usually does not make any assumption about models. Instead, the problem is typically tackled using the state-action value function, or the Q-function defined as

$$q_\pi(s, a) = \lim_{h\to\infty} \sum_{t=0}^{h} E[\gamma^t r(s_t, \pi(s_t))|s_0 = s, a_0 = a]. \tag{5}$$

The Bellman optimality equation characterizes the optimal q_* when an optimal control policy $\pi^\star$ is achieved [11]:

$$q_*(s) = \sum_{s',r} p(s', r|s, a)[r + \gamma \max_{a'} q_*(s', a')]. \tag{6}$$

Here q_* is defined similarly as v_*.

3.2. State Space

In the WT operation problem, the state space includes *the physical states* that can be measured and controlled and *the exogenous states* that can be measured but are uncontrollable.

The physical control variables collected from the wind turbine typically include rotor speed, rotor torque, generator speed, generator torque, and blade pitch angle. Based on the physical mechanisms of wind turbines discussed in Section 2, we take the pitch angle and generator torque as control variables. Most literature chose the variable of generator torque, which can then control the rotor speed, the rotor torque, and the power production. The exogenous uncontrollable variables are mainly the wind speed and wind direction, which influence the power output.

Therefore the state of our MDP model is $\boldsymbol{s} = (s_1, s_2, s_3, s_4)$ where s_1 denotes the pitch angle measured in degrees, s_2 is the generator torque in Nm, s_3 is the average wind speed in *m/s*, and s_4 is the corrected absolute wind direction in degrees.

3.3. Action Space

For the two main control variables, the pitch angle and generator torque, the action space is defined by the change of pitch angle and torque, $a = (a_1, a_2)$, where a_1 and a_2 represent the change of pitch angle in degrees and the change of torque in Nms, respectively.

3.4. Reward Functions

MPPT aims to maximize the power output in the long term. Therefore, for each action, the reward is measured by the average power output generated in the next epoch. The reward function maps the state space to a real value.

For wind turbine control, it is not practical to change the generator torque constantly in time, since it can cause the machine subject to unnecessary stress. Therefore, we consider that the control action is taken at discrete time, e.g., every hour, every day.

3.5. Model-based Off-Policy Reinforcement Learning for Maximizing Average Reward

As both state space and action space are continuous, we consider using function approximation in reinforcement learning, where the Q-function is approximated. One of the most straightforward algorithms to handle this scenario is

the fitted Q iteration algorithm [15]. In the fitted Q iteration, the approximation function can be in any form and is an approximation of the state-action value function $Q_a(s, a)$: $S \times A \rightarrow R$.

The algorithm assumes a greedy policy, given a training set (s_i, a_i, r_i, s'_i) for $i = 1,\ldots, N$. In each iteration, we first estimate the Q function for each training quadruple from the current approximation according to Bellman equation

$$q_i \leftarrow r_i + \gamma \max_{a \in A} Qa(s'_i, a). \tag{7}$$

Then we update the function approximation with the new estimated Q function value for the training set

$$\theta \leftarrow \min_{\theta} \sum_{i=1}^{N} L(Qa(s_i, a_i; \theta), q_i), \tag{8}$$

where γ is the discount factor, θ denotes the parameters of the Q function approximation, and L is a loss function such as the squared error loss.

The algorithm is able to find the policy that maximizes $J_{\infty}^{\pi}(s)$ in (2) [15]. In the discounted case with infinite horizon, the algorithm converges to the optimal policy that is a stationary point of the Bellman equation for qualified models. The detailed conditions are described in [16].

Our aim is to maximize the long-term average reward without the existence of the discount factor, which is defined as [16]

$$R_{\infty}^{\pi}(s) = \lim_{h \to \infty} \frac{1}{h} \sum_{t=0}^{h} E[\gamma^t r(s_t, \pi(s_t)) | s_0 = s]. \tag{9}$$

The relationship between $J_{\infty}^{\pi}(s)$ and $R_{\infty}^{\pi}(s)$ is given by [12]

$$J_{\infty}^{\pi}(s) = \frac{1}{1-\gamma} R_{\infty}^{\pi}(s). \tag{10}$$

Hence, we can do the fitted Q iteration according to (10) to reach the optimal policy and calculate the optimal long-term average reward.

4. Empirical and Simulation Studies

4.1. Data Description

The wind turbine considered in this research is Senvion MM82 2.05MW wind turbine [16]. From its technical specifications, we are able to obtain $s_3 \in (3.5, 25)$ that indicates the wind speed to lie between the cut-in wind speed and the cut-out wind speed [17]. However, the ranges for other variables are not specified in the technical specifications and can be estimated from historical data. The estimated ranges are $s_1 \in [-1, 40]$, $s_2 \in [0, 6000]$ and $s_4 \in [0, 360]$. The ranges of action a_1 and a_2 are specified according to the current state to make sure that the next state lies in the normal range.

The data we analyze in this research was collected at La Haute Borne wind farm located in Meuse, France by ENGIE [17]. The SCADA data were obtained from wind turbines R80721, R80711, R80790, and R80736 since 2013. Due to the storage and IO limit, we are able to access the maximum, minimum, and average values of signals in 10-minute intervals, although the original data were collected at a higher frequency.

In wind turbine operations, it is not practical to make real-time decisions that constantly change the control variable, due to the relatively slow reaction of the electronic and mechanic components of wind turbines. In this research, we take the average value in a one-hour interval as the observed value for each variable at each decision epoch.

4.2. Model Selection

To explore the predictability of different models, we first use the state-action pairs as predictors to estimate the power output at the next time step. The mean absolute error and the maximum absolute error are used to measure the accuracy of the models, presented in Table 1.

Table 1: Mean absolute error and maximum absolute error of predictors

Model	Mean Absolute Error	Maximum Absolute Error
Linear	9.14	50.54
Quadratic	6.89	69.07
Random Forest	6.01	101.89

We compare our results from different models to the results in [8]. All the models in Table 1 provide fine estimators in terms of the mean absolute error. Hence, we decide to choose the linear model for its simplicity and the lowest overfitting according to the maximum absolute error.

4.3. Fitted-Q Iteration Results

As the sum of absolute change in q_i, the convergence threshold does not exceed 1,000. When $\gamma = 0.6$, the estimated long-term average reward is 1,145kW; and when $\gamma = 0.8$, the estimated long-term average reward is 1,140kW, which is about 730kW on average per step higher than the observed average reward. The optimal action is always maximizing the pitch angle (40) and generator torque (6000), which is a natural optimal condition for a linear model. The impractical action values could be the result of our training samples that were taken from the real operation process with a strong positive correlation among wind speed, generator torque, and power output.

To avoid the strong correlation among the variables, we conducted an experiment in which only the pitch angle is taken as the control variable. In this experiment, the quadratic regression was used for function approximation, while other parameters remain the same. Figure 1 shows how the optimal pitch angle changes with the wind direction, as the results of optimal control for maximizing the long-term average reward to be 507kW. In Figure 1, the x-axis denotes the decision epochs and the y-axis is the degrees of angle.

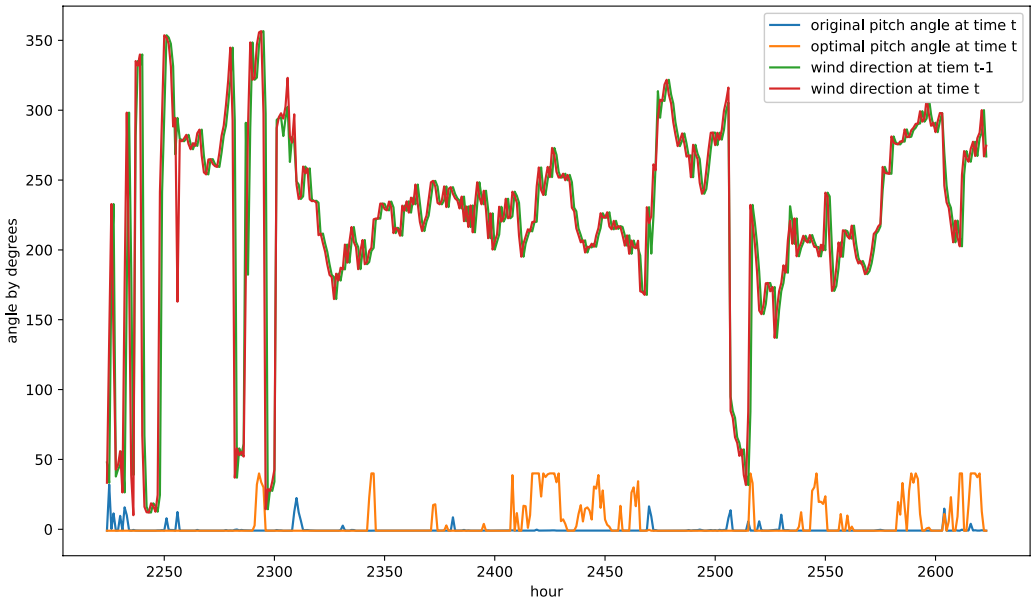

Figure 1: A snapshot of the changes in the optimal pitch angle along with the wind direction

5. Discussion and Conclusions

In this research, we maximize the power output of wind turbines under the stochastic wind profile by formulating the problem as a continuous state space, continuous action space Markov decision process. The curse of dimensionality

of DP methods is overcome by utilizing the function approximation in reinforcement learning, since exact methods become infeasible for the large MDPs. In the computational studies using real data, we conducted preliminary research by applying the fitted-Q iteration algorithm to the MPPT task of wind turbines. With the MDP formulation, the consecutive decision-making and control delay are embedded in the model and we can further generalize it to online version with policy gradient techniques [16]. Instead of the linear model, more complex models can be used such as non-linear models, supervised learning, or a neural network method.

Moreover, the optimization over Q_a can lead to impractical action values, which could be the result of our training samples that were taken from the real operation process with a strong positive correlation among wind speed, generator torque, and power output. This hypothesis has been validated in our experiment by excluding the generator torque in control variables. To incorporate a dependent control variable such as the generator torque, we will carry out data augmentation over data points with large generator torque values. In this situation, from a data science point of view, we should first improve the model for Q_a. Modern artificial neural network models could be promising for function approximation.

References

[1] Eduardo José Novaes Menezes, Alex Maurício Araújo, and Nadége Sophie Bouchonneau da Silva. A review on wind turbine control and its associated methods. Journal of Cleaner Production, 174:945–953, 2018.

[2] Karin Ohlenforst, Steve Sawyer, Alastair Dutton, Ben Backwell, Ramon Fiestas, Joyce Lee, Liming Qiao, Feng Zhao, and Naveen Balachandran. GLOBAL WIND REPORT 2018. 04 2019.

[3] Lucy Y Pao and Kathryn E Johnson. Control of wind turbines. IEEE Control Systems Magazine, 31(2):44–62, 2011.

[4] U.S. Energy Information Administration. Electric Power Monthly. https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_6_a. 2019.

[5] Jackson G Njiri and Dirk Soeffker. State-of-the-art in wind turbine control: Trends and challenges. Renewable and Sustainable Energy Reviews, 60:377–393, 2016.

[6] Debashisha Jena and Saravanakumar Rajendran. A review of estimation of effective wind speed based control of wind turbines. Renewable and Sustainable Energy Reviews, 43:1046–1062, 2015.

[7] Jannis Tautz-Weinert and Simon J Watson. Using scada data for wind turbine condition monitoring–a review. IET Renewable Power Generation, 11(4):382–394, 2016.

[8] Andrew Kusiak and Haiyang Zheng. Optimization of wind turbine energy and power factor with an evolutionary computation algorithm. Energy, 35(3):1324–1332, 2010.

[9] Andrew Kusiak, Wenyan Li, and Zhe Song. Dynamic control of wind turbines. Renewable Energy, 35(2):456–463, 2010.

[10] Henk C.Tijms. A First Course in Stochastic Models. John Wiley & Sons, Ltd, 2004.

[11] C. Wei, Z. Zhang, W. Qiao, and L. Qu. Reinforcement-learning-based intelligent maximum power point tracking control for wind energy conversion systems. IEEE Transactions on Industrial Electronics, 62(10):6360–6370, Oct 2015.

[12] Richard S. Sutton. Learning to predict by the methods of temporal differences. Machine Learning, 3(1):9–44, Aug 1988.

[13] Majid A Abdullah, AHM Yatim, Chee Wei Tan, and Rahman Saidur. A review of maximum power point tracking algorithms for wind energy systems. Renewable and Sustainable Energy Reviews, 16(5):3220–3227, 2012.

[14] Martin L Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, 2014.

[15] Damien Ernst, Pierre Geurts, and Louis Wehenkel. Tree-based batch mode reinforcement learning. Journal of Machine Learning Research, 6(Apr):503–556, 2005.

[16] Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction. MIT press, 2018.

[17] Senvion S.A. MM82 wind turbine 2MW. https://www.senvion.com/global/en/products-services/wind-turbines/mm/mm82/, 2019.

[18] ENGIE. Engie opendata. https://opendata-renewables.engie.com/explore/index, 2019.