pp.45:50

ISSN: 2251-924
EISSN: 2345-6221

Using Reinforcement Learning to Make Smart Energy Storage Source in Microgrid

Sadegh Etemad [1], Nasser Mozayani[2]

School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran,Email: S_Etemad@Comp.IUST.ac.ir
School of Computer Engineering, , Iran University of Science and Technology, Tehran, Iran,Email: Mozayani@IUST.ac.ir

Abstract

The use of renewable energy in power generation and sudden changes in load and fault in power transmission lines may cause a voltage drop in the system and challenge the reliability of the system. One way to compensate the changing nature of renewable energies in the short term without the need to disconnect loads or turn on other plants, is the use of renewable energy storage. The use of energy storage improved electrical stability, power quality and improve the peak power load. In this paper, we have used the reinforcement learning to present an optimal method for charge and discharge the consumer battery. In this way the uncertainty of production due to the random nature of wind energy is improved. Simulation results indicate not only the use of renewable energy and battery is successfully enhanced but also the cost of annual payments and peak consumption times is reduced.

Keywords: Rreinforcement learning; Microgrid; Renewabl energy; Battery; Q learning; Intelligence agent

1. Introduction

Nowadays contaminant gases such as carbon dioxide, nitrous oxide produced by the burning of fossil fuels are one of the concerns of governments and societies. The release of these gases leads to climate change and warmer unusual atmosphere, irreparable harm to human health. Therefore, reducing the production of such gases is essential for all industries such as the power industry. Reducing wasting of energy in different ways such as decreasing energy waste in power transmission lines, modifying the pattern of electricity consumption and use of renewable energy instead of fossil fuels can reduce the emission of these gases into the air.

Widespread use of renewable energy in power generation requires changes in the power system. This is because these resources are uncertain in electric power generation due to their random nature. Therefore, it is better to use store energy sources in order to provide energy during peak times. During low load times, energy generated by wind turbines with low contamination levels, can be stored and supplied during the peak load times, rather than using the production units with high pollution. To this end, we need an intelligent algorithm for charging and discharging of the energy storage.

1. Related Work

In coming years, consumer wants smart machines and expects machines to think and operate autonomously and optimally. Reference [1] energy management of microgrids using fuzzy logic discussed. References [2] and [3] using genetic algorithm to smart management of microgrids. Energy management of hybrid renewable energy generation using constrained optimization was proposed in [4]. Classical and heuristic algorithms for energy management of microgrids in [5] and [6] also expert system in order to energy management microgrids in [7] proposed. In all this method, the concept of learn is not emphasized and interactive learning is not considered.

In [8] to overcome the unpredictability of the electricity market and a rapid response to this uncertainty have provided a framework based on reinforcement learning agent. References [9] distributed control microgrid have done by Using the framework of the reinforcement learning multi-agent systems. In [10] offered A fuzzy Q learning method based on genetic algorithms for energy management in smart grids and in [11] offer Smart microgrid electricity flow management using multi-agent reinforcement learning.

In this paper, Q-learning, one of the reinforcement learning methods, to manage energy storage has been used. This method is able to use the past experience of the intelligent battery agent and choose the appropriate action between charge and discharge actions for next one hours.

The rest of the paper is organized as follow. Section three present modelling of the wind microgrid and the details of component in it. In section four reinforcement learning and Q-learning method has been explained. Section five provides a model for smart battery agent. In section six case study and simulation result are analyzed. At the end, in section seven is conclusion.

2. Wind Microgrid Model

Microgrid is set of distributed generation of electricity that form small-scale and low voltage grid to feed electrical and thermal loads. The simplified model of microgrid is depicted in Fig. 1. According to the figure, microgrid contains three components.

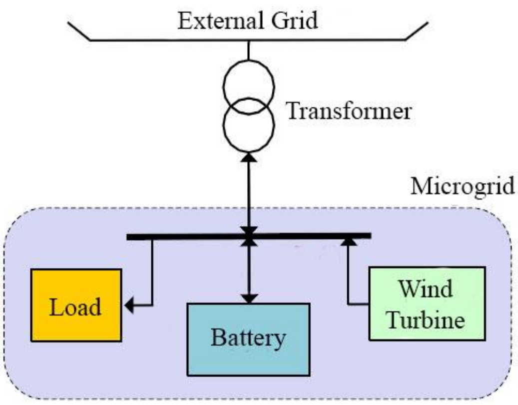


Fig. 1. A simplified model of microgrid [12]

A. Wind turbine

Wind speed, air density, the radius of the wind turbine blades has a significant impact on the output power of wind turbine. In this paper, power output from the wind turbine at time step t from (1), in [13] presented, are calculated. Where v_t is the working wind speed (m/s) at time step t of 1h, v^{ci}, v^r and v^{co} are the cut-in, rated and cut-off wind speeds (m/s), respectively, and P^r is the rated power of the wind turbine (W).

$$P_t^{wt} = \begin{cases} 0 & v < v_{ci} \\ P^r.(\frac{v_t - v^{ci}}{v^r - v^{ci}}).\Delta t & v_{ci} < v < v_r \\ P^r.\Delta t & v_r < v < v_{co} \\ 0 & v > v_{co} \end{cases} \quad (1)$$

B. Battery

Various technologies use energy from batteries as storage, but the batteries are electrochemical cells. Now the batteries are the best options for storing electrical energy in small and medium levels. A simplified model of battery storage dynamics is adopted by implementing a discrete time system for the power flow dynamics over time step interval Δt is shown in (2) [14],

$$R_t = R_{t-1} + R_t^{Stor,Charge} - R_t^{Stor,Discharge} \quad (2)$$

Where R_t and R_{t-1} are the levels of the energy stored in the battery at time t and t-1 (Wh), $R_t^{Stor,Discharge}$ and $R_t^{Stor,Charge}$ are the power flows over time step interval Δt between battery and consumer, wind generator and battery (Wh), respectively.

3. Reinforcement learning

In machine learning, there are issues that are too small and inadequate resources to solve the problem therefore possibility of using supervised learning algorithms does not exist. In such circumstances reinforcement learning can be used that learning will be done by experience.

Learning about what action needs to be done to maximize numerical reward signal, in each case, called reinforcement learning. In this method the learner about what to do, such as common forms of machine learning is not knowledge, but the agent must try to discover what practical actions ahead at what the situation is the most rewarding. Actions not only effect on present reward but also states and future reward affected. The Search-based trial and error and delayed rewards of the most distinguishing features are reinforcement learning

A. Reinforcement Learning model

In the standard reinforcement learning model, an agent is connected to its environment via perception and action, as depicted in *Fig. 2*. On each step of interaction, the agent receives as input some indication of the current state of the environment; the agent then chooses an action to generate output. The action changes the state of the environment and the value of state transition is communicated to the agent through a scalar reinforcement signal. The agent's behavior should choose actions that tend to increase the long-run sum of values of the reinforcement signal, it can learn to do this over time by systematic trial and error, guided by a wide variety of algorithms [15].

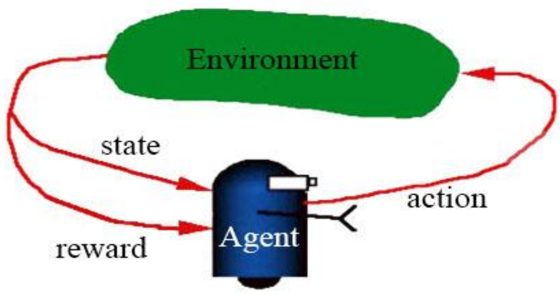


Fig. 2. Standard model of reinforcement learning

Formally, this model consist of three main set :
- A discrete set of environment state, S
- A discrete set of agent actions, A
- A set of scalar reinforcement signals; typically, {0,1}, or the real numbers.

B. Q learning

Standard reinforcement learning or Q learning a learning algorithm that can be based on delayed reward and reward function learn optimal policy. The purpose of the policy mapping the set of observed states to set of actions. The output of this algorithm is amount of Q table where amount of dual action-state stored. Q learning process include following step:
- Initialized Q table with zero
- Received current state of the environment, s
- Repeat the following loop until the ending condition
- Choose action a through one of the following two method
- Randomized (exploration)
- Based on current policy, Q table (exploitation)
- Receive reward from environment, r
- Receive next state of environment, s'
- Update value of Q table based on (3),

$$Q(s_t,a_t) = Q(s_t,a_t)+\alpha.(r(s_t,a_t) + \gamma.Max_{a'}(s'_t,a'_t) - Q(s_t,a_t)) \quad (3)$$

Where $Q(s_t,a_t)$ is the value of dual action-state in Q table at time step t, α is the learning rate that value is between 0 and 1 that influences the speed of convergence to final Q (Q*), $r(s_t,a_t)$ is reward function for state s and action a at time step t, γ is the discount factor that value is between 0 and 1 that contributes to determining the value of future reward function and a'_t is the action at next time step t'[16].

The algorithm starts each time with an initial state s and performing a sequence of actions. It receives reward to reach goal state. Usually in goal state, state of agent by performing any action does not change and will not receive any reward from environment.

4. Smart Battery Model for Consumer

In order to be able to have an intelligent model for a battery we need mapping between the intelligent battery modeling and reinforcement learning model we have. According to what was said reinforcement learning model three sets of actions and rewards states will be required. In this section we consider this map.

A. Environment and the Actions of Battery

Available wind power output and consumer load are the two variables defining the dynamic environment of consumer. The values of these two variables at time step t define the system state. Equation (4) is shows the environment state.

$$S_t = (L_t, P_t^{wt}) \quad (4)$$

Where L_t , is the consumer load at time step t, P_t^{wt} is available wind power output. Furthermore, the level of battery charge R_t at time step t must also be considered. The actions that were taken for battery are charging and discharging. In (5) a_{ch} means for charging the battery via wind turbines and a_{dis} means the battery discharge by the consumer is in microgrid.

$$A_t = \{a_{dis}, a_{ch}\} \quad (5)$$

B. Reward Function

Since the battery at any time t can only perform one action, acts simultaneously charging and discharging are not possible. Accordingly, the reward function through the (6) is calculated [17].

$$r_t(a_t) = \begin{cases} \frac{P_t^{wt}}{L_t}.(L_t - R_t^{Stor,Discharge}) & a_t = a_{dis} \\ K.(P_t^{wt} - R_t^{Stor,chrage}) & a_t = a_{ch}, P_t^{wt} > 0 \\ 0 & a_t = a_{ch}, P_t^{wt} = 0 \end{cases} \quad (6)$$

Where K is a weight coefficient for the reward function when action is a_{ch}, $R_t^{Stor,Discharge}$ and $R_t^{Stor,chrage}$ are the amounts of electricity (Wh) that the battery is capable of discharging and charging during 1 h, respectively. Observe that by increasing the value of the weight coefficient k the value of reward obtained after action a_{ch} increases in comparison with the reward obtained after action a_{dis}, which makes action a_{ch} more likely to be performed.

C. Evaluation of Battery Actions

In this paper we aim to find the optimal action at time t for the next hour (t+1) that is charged or discharged. For that, first the Q table values initialized by zero. After percept state of the environment will use exploration and exploitation to choose appropriate action then rewarding value calculated and value of Q table will be updated. This will continue until all Q values converge to maximum Q-value (Q*). Practical, the highest Q* in the current case have been chosen to act as appropriate for the operating battery.

5. The Results of Simulation and Analysis

In this section summary of the simulation results for proposed reinforcement learning algorithm is presented. Because of the high computational complexity of continuous reinforcement learning algorithm, discrete Q-learning algorithm have been used to simulate. Accordingly, for each of the variables power output of wind turbine, consumer load and battery charge level six discrete state is considered. Wind turbine parameters considered is visible in table I.

Table.1.
Characteristic of wind turbine

Parameter	P^r	v^{ci}	v^r	v^{co}
Values	6000 W	3 m/s	12 m/s	20 m/s

Using equation (1) and table I power produced by wind turbines discrete values are {0, 1200, 2400, 3600, 4800, 6000}. Amounts intended for the consumption of electric systems through independent farmers in Canada, the Province of Ontario is obtained. This operator provides data of consumer load, power output and price per watt hour since 2002 to the present. Also prediction this parameter for 24 hours ahead is available for researcher in this institute. Values considered for consumer load is {2000, 2800, 3600, 4400, 5200, 6000}. Finally, the values considered for battery charge level includes {0, 1000, 2000, 3000, 4000, 5000}. We have assumed which each charge or discharge action cause the battery level 1000 (Wh) changed in one-hour period. Maximum capacity of the battery to 5000 (Wh) and minimum capacity of the battery is 0. Parameters α and γ was considered 0.8 in implementation. Also value of K is 6. This value through empirical experiment calculated

We used data from Canada at 2009 to the end of 2014 year to learnt Q-learning algorithm. *Fig. 3* shows amounts of an annual sum of electricity production and consumption in the years 2009 to 2014.

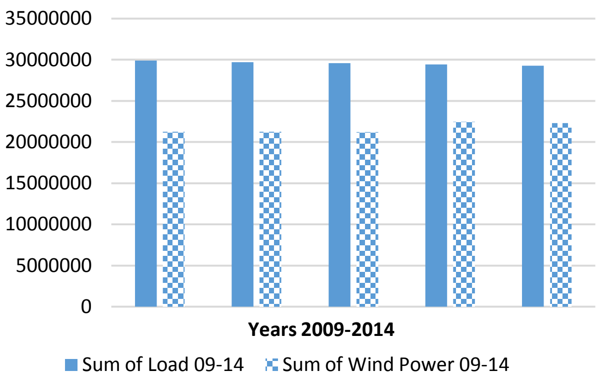


Fig. 3. Annual sum of electricity production and consumption

The purpose of learning is finding final Q-table. Fig. 4 show test data is considered from the first 7 months of 2015 year.

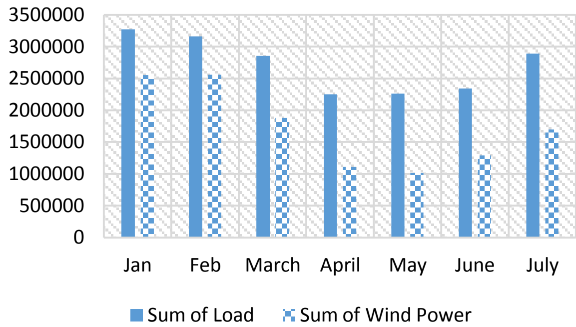


Fig. 4. Monthly sum of electricity production and consumption

6. Evaluating the Performance of Microgrid

The proposed algorithm gives two major goals:

– Increasing the utilization rate of the battery during high electricity demand (so as to decrease the electricity purchase from the external grid)

– Increasing the utilization rate of the wind turbine for local use (so as to increase the consumer independence from the external grid)

To evaluate the proposed algorithm defines three parameters that they indicate the aim is spoken. Equation (7),(8) and (9) show this parameters[14].

$$V_0 = \frac{\sum R_t^{Stor,Discharge}}{\sum L_t} \tag{7}$$

$$V_1 = \frac{\sum R_t^{Stor,charge}}{\sum P_t^{wt}} \tag{8}$$

$$E = (\sum L_t - \sum R_t^{Stor,Discharge}).pr_t \tag{9}$$

Where in (9), pr_t is the price of electricity is produced by wind turbines at time t. Because the wind turbine costs related to the cost of setting up and maintenance, the cost of electricity production after the setup of a fixed amount of $ 1 per watt hours is considered. V_0 in (7) represents the rate of use of the battery power to the total amount of load annually. Increase this parameter represents an increase in the rate of battery used. V1 in (8) show electricity rates used by the battery in order to charge to the total amount of power produced by the wind turbine in a year that increase it means increased use of renewable energy in here wind turbine. The parameter E represents the annual cost of buying electricity from an external grid. Unlike the two parameters, reduce this parameter equal to reduce the annual cost.

To evaluate the improvements of the performance indicators achieved by the application of the reinforcement learning for microgrid management under stochastic wind speed conditions, 20 independent simulation runs are executed. *Fig. 5* Represents the V_0 in seven month test data. As you can see this parameter during 7 months the trend has been upward. This chart shows the increase use of battery in microgrid.

Figure 6 Represents the V_1 in seven months test data. Ascending of this graph is also a sign of the increasing use of wind energy.

Figure 7 shows the values E for 7 months. The vertical axis represents the price is in dollars per Wh. Strictly descending is sign the cost reduction per year.

In order to test the learning process, randomly two days data from the first 7 months of 2015 year related to province of Ontario selected and studied the performance of the proposed algorithm on it. Figure 8 shows the two random day and the charge or discharge the battery. The initial battery charge is zero.

As shown in Figure 8(a) you can see a Smart Agent do the best action for the next time t + 1. Figure 8(b), except for two hours of 6 and 7 that would make charging at other times the best sequence of actions is chosen.

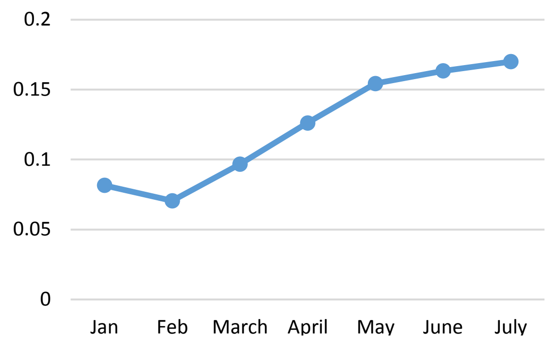


Fig. 5. V_0 average value for 7 months

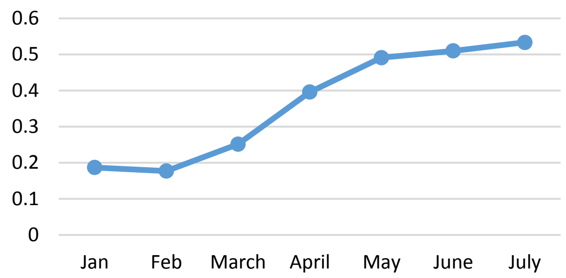


Fig. 6. V_1 average value for 7 months

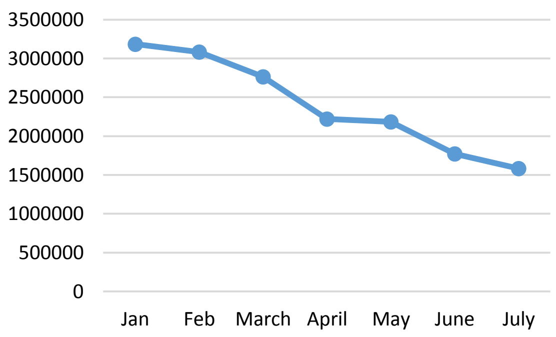
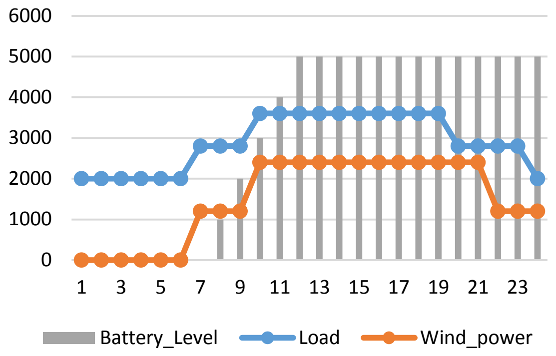


Fig. 7. E average value for 7 months

7. Conclusion

In this article we use reinforcement learning method to provide intelligent battery energy storage source in the microgrid. The intelligent battery selects appropriate action from charge or discharge based on the amount of consumer load, wind turbine production and the battery power level. The aim of selecting the best action at time t is the performing of the action at the time t+1.

Due to the uncertain nature of the renewable energies, using the storage sources of energy is recommended. The purposed method covered this uncertainty and can be used energy stored in the battery when the wind power generation does not exist. The results of the simulation indicate that the proposed method was to improve the state of uncertainty. Also in this method, Smart Battery operating without a supervisor and autonomously based on their experiences choose best action. In addition, when power generation is lower than consumer load within the microgrid consumer can use electricity stored in the battery instead of buying electricity at high prices from the external grid. This action in addition to reduce costs, reduce peak load electricity and increases reliability, moves toward ultimate goal that is no blackout.



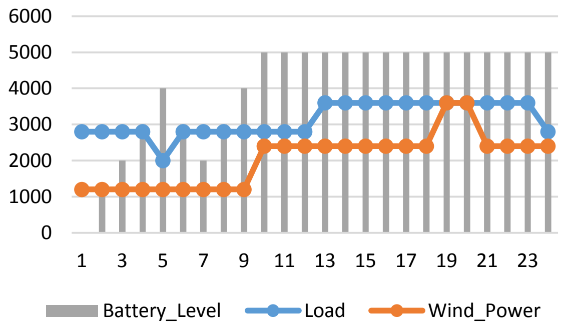


Fig. 8. Performance of battery Agent (charging, discharging) for 24 hours two days randomly

References

[1] N. Hatziargyriou "Microgrid and energy management," European transaction on electrical power, pp. 1139-1145, December 2010.

[2] P. Reddy and M. Veloso "Strategy learning for autonomous agents in smart grid markets," Twenty-second international joint conference on artificial intelligence, pp. 1146-1151, December 2005

[3] Chen, S. Duan, T. Cai, B. Liu and G. Hu "Smart energy management system for optimal microgrid economic operation," Renewable Power Generation, vol. 5, pp. 258-267, 2011.

[4] Mohamed and N. Koivo "System modelling and online optimal management of microgrid with battery storage," International journal on Electrical Power and Energy Systems, vol. 32, pp. 398-407, 2010.

[5] C. Colson, M. Nehrir and A.Pourmousavi "Toward real-time microgrid power management using computational intelligence methods," Power and Energy Society General Meeting, vol. 8, pp. 1-8,July 2010.

[6] M. Abdirahman, M. Abdilahi, W. Mustafa, G. Aliyu and J. Usman "Autonomous integrated microgrid system," International journal of Education and Research, vol. 2, pp. 77-82, January 2014.

A. Chaouachi, M. Rashad, M. Kamel, R. Andoulsi and K. Nagasaka "Multiobjective intelligent energy management for a microgrid," IEEE Transactions Industrial Electronics, vol. 60, pp. 1688-1699, 2013.

[7] Y. Guo, A. Zeman, and R. Li, "A Reinforcement Learning Approach to Setting Multi-Objective Goals for Energy Demand Management," Int. J. Agent Technol. Syst., vol. 1, no. 2, pp. 55–70, 2009.

A. L. Dimeas and N. D. Hatziargyriou, "Multi-agent reinforcement learning for microgrids," Power Energy Soc. Gen. Meet. 2010 IEEE, pp. 1–8, 2010.

[8] X. Li, C. Zang, W. Liu, P. Zeng, and H. Yu, "Metropolis Criterion Based Fuzzy Q-Learning Energy Management for Smart Grids," Control Conf., vol. 10, no. 8, pp. 1956–1962, 2012.

[9] Lauri, G. Basso, and J. Zhu, "Managing Power Flows in Microgrids Using Multi-Agent Reinforcement Learning," Agent Technol. Energy Syst., 2013.

[10] Changbin, L. Shanna, L. Zhengxi, W. Xin and L. Sun "Energy coordinative optimization of wind-storage-load microgrids based on short-term prediction," Energies journal, vol. 8, pp. 1505-1528, April 2015.

[11] S. Roy, "Market constrained optimal planning for wind energy conversion system over multiple installation sites,"Energy Conversion IEEE Transaction, vol. 17, pp. 124-129, August 2002.

[12] E. Kuznetsova, L. Yan-Fu, C. Ruiz, E. Zio and G. Ault "Reinforcement learning for microgrid energy management," Energies journal, vol. 59, pp. 133-146, July 2013.

[13] L. Kaelbling, M. Littman and A. Moore "Reinforcement learning: A Survey," Journal of Artificial Intelligence Research, vol. 4, pp. 237-285, July 1996.

[14] R. S. Sutton and A. G. Barto, "Reinforcement learning:An Introduction," London, England, MIT press,2005.

[15] R. Leo, R. S. Milton and S. Sibi "Reinforcement learning for optimal energy management of a solar microgrid," IEEE Global Humanitarian Technology Confrence, India, September 2014.