

Research of Q-learning Algorithm of Multi-agent in Micro-grid Control System Based on Probability

Li Xiangna^{1, a}, Yi Nan^{2, b}, Li Mengao^{3, c} and Xu Weifeng^{3, d}

¹Beijing GuoDianTong Network Technology Co., Ltd, Beijing, China

²North China Electric Power University, Baoding, China

³China Shipbuilding Industry System Engineering Research Institute, Beijing, China

^alxn1983@126.com, ^btianqing591@sohu.com, ^csoutheast1983@126.com, ^dweifengxu@163.com,

Keywords: Micro grid control; Multi-agent structure; Greedy Algorithm.

Abstract. Aiming at the problems existing in Multi Agent Q learning algorithm in micro-grid control systems, this article put forward a Q-learning algorithm of Multi-agent in Micro-grid control system based on probability. This algorithm introduces the theory of probability into greedy strategy, so that the optimal action will be chosen. At the same time in order to take the influence of mutual problems between agents and the historical information of every agent to the action selecting problem into consideration, the historical information is added as a parameter during getting the prime function of finding the ideal value. In order to test my algorithm, this article tests the algorithm in emulated micro-grid control system. It appears that , this algorithm can restore the power to stable state rapidly, when the power of the micro-grid changes frequently.

Introduction

Currently, fossil energy around the world has been running out, development and utilization of renewable energy sources has become a new trend. Because electricity possesses many features such as clean, efficient, transforming new energy and renewable energy into electrical energy is an effective way of energy utilization. Micro-grid technology is a technology using the new green renewable energy to generate electricity. Micro-grid is distributed, it includes a few complex electrical equipment, if it the operation of these devices in the micro network systems can be effectively controlled, micro-network will be more perfect. In such a trend is a more effective energy management system - Multi-Agent System is acquired to ensure the reliability and effectiveness of the power system, and the core issue of Multi-Agent Systems is to study its learning algorithm^[1, 2]. Research in the past 10 years for Multi-Agent Systems learning has become a topic of concern in Multi Agent Systems learning, in the field of multi-agent system learning, the technology of learning-reinforcement attracted the attention of many scholars, this is because learning-reinforcement does not require environmental model, and allows multiple Agents to take an interactive approach to learning^[3].

This paper introduces the micro-grid control systems in the workflow of each Agent, and then improve the multi-Agent Q-learning algorithm, and finally regard IEEE 9-bus system as a micro-grid system and run an emulation in this system, the simulation results show that when micro-grid power fluctuations occur quickly the algorithm can restore the power to a stable state fast. This reflects the feasibility of the algorithm.

Multi-Agent Workflow in Micro-grid Control System

Micro-grid control system is constituted by a plurality of agent. The loaded agent monitors each load and evaluates the performance of each load, at the same time diagnosis errors. The loaded Agent will be informed of the unusual information. Micro-grid agent controls the decision making according to the information, and commands the micro-source agent to solve the problem through collaboration^[4].

Improved Multi-agent Q learning algorithm in micro-grid control systems

Micro-grid control system is a Multi Agent system that works in a dynamically changing environment, and there is no complete prior knowledge and prior model. Therefore a leaning algorithm, which is applicable to micro-grid controlling, should have the following characteristics: processing large amounts of data and variables, adapting to a variety of changing systems, the algorithm should be operated in a macro-environment. Since the environment is random, each action is uncertain.

However, right now some problems still exist in the Q learning. These include the action-selection, multi-Agent interaction issues, and extraordinary complexity of the spatial state of the actions of the problem^[5]. In the respect of action selection, greedy strategy is more commonly used. The applying of greedy strategy will make the value of Q reach the climax very quickly, and this often make the Agent reach the local optimality at an early stage of the calculation, and then converge at the local excellent peak. This method can easily miss more ideal possible action.

Q learning is based on a single-step process of learning, it's learning process just use current information to update the Q matrix, which results in slower computation algorithm, because there is always the need for a relatively longer time to search for the optimal operation. The circumstance in which intelligent algorithm is used, is based on a real-world-based model. In a mixed complex system there are not only the interaction between each Agent and the environment , but also the interaction between the Agents. Therefore single step Q learning is no longer applicable to multi- Agent system.

Based on analysis above, the multi-Agent learning method is a interactive and collaborative learning methods. In the multi-Agent system, environment changes under the combined effects of multiple Agents , each Agent action need to, while performing, interact with other Agent and conceive the action selection of other actions. This is possible in some small-scale systems. Once the number of Agents, or the optical actions under different circumstances grows, then the spatial state of the agents will grow exponentially. Theoretically this kind of interaction is possible. But when it comes to the reality, the capability of computer can hardly implement the calculation. There for the extraordinary complexity of spatial state of the actions reduces the speed of Multi-agent learning.

improved multi-Agent Q learning algorithm theory

To address the problems in Q algorithm, this paper proposes an improved multi-Agent Q learning algorithm, the method works as follows:

(1). In greedy strategy^[7], we no longer use the traditional pure high Q value to determines the method of action-selecting, but select the action according to the probability value .We Suppose a Agent obtained, by performing a certain action, the next state. If the Q value of the next action is relatively large and the probability that this action is selected is relatively large, the action will be selected. The possibility that an action will be selected, will be calculated as follows:

$$p(s|a_i) = \frac{k^{Q(s,a_j)}}{\sum_j k^{Q(s,a_j)}} \quad (1)$$

Where in k ($0 < k \leq 1$) is a constant. Selecting the action according to probability sufficiently avoid the local optimization proble.

(2).We also added to the algorithm the historical information of other Agents, Qfunction becomes $Q(s,h,a)$ representing the bonus value of the information. At the same time using indirect interactive reinforcement method find the optima solution path through backtracking the previous action selection. Each Agent will leave ,after each learning, "marks" in the environment, the more significant the marks are, there will be more agents pass the position, and the more likely this position is part of the optimal solution path. We represent the significance degree of this "mark" with M_{ij} . The updating formula is:

$$m_{ij}^t = \sum_{k=1}^n \Delta m_k^t \quad (k = 1, 2, \mathbf{K}, n) \quad (2)$$

$$M_{ij} = (1 - \mathbf{I})M_{ij}^t + \mathbf{I}M_{ij}^t \quad (3)$$

Wherein Δm_k^t is the significance degree of Agent K at coordinate(i,j) (if Agent k ($k = 1, 2, \mathbf{K}, n$) at time spot t passes the coordinate (i, j) , we set $\Delta m_k^t = 1$), m_{ij}^t is the significance degree of mark of the Agent at coordinate $((i, j))$. M_{ij}^t is the historical “mark” significance degree at the coordinate (i, j) . \mathbf{I} is the persistence of the “mark”. $1 - \mathbf{I}$ is the degree of attenuation.

Based on the analysis above, the improved Q learning algorithm can be expressed as:

$$Q_t^i(s_t, h_{t-1}, a_t) = (1 - a)Q_{t-1}^i(s_t, h_{t-1}, a_t) + a[r_t^i + M_{ij} + bV^i(s_{t+1}, h_t)] \quad (4)$$

wherein, $V^i(s_{t+1}, h_t) = \max_{a_{t+1}^i} \sum_{a_{t+1}^{-i}} P_r((a_{t+1}^i, a_{t+1}^{-i})|h_t, s_{t+1})Q_{t-1}^i(s_{t+1}, h_t, a_{t+1})$. Historical information is added, so that the information of other Agents can also be considered during the decision-making. The strategy is:

$$\Pi_i(s) = \arg \max_{a_{t+1}^i} \sum_{a_{t+1}^{-i}} P_r((a_{t+1}^i, a_{t+1}^{-i})|h_t, s_{t+1})[r_{s_{t+1}} + M_{ij} + gQ(s_t, a_t)] \quad (5)$$

Improved multi-Agent Q learning algorithm iteration step

Initializing the state set and action set of each Agent, we set the learning rate a And discount rate b . We set $k=1$, $\mathbf{I}=0.2$ and $Q(s, h, a) = \frac{1}{n}$, where n represents the number of executable action a under the state s . Its iterative process is shown in Figure1.

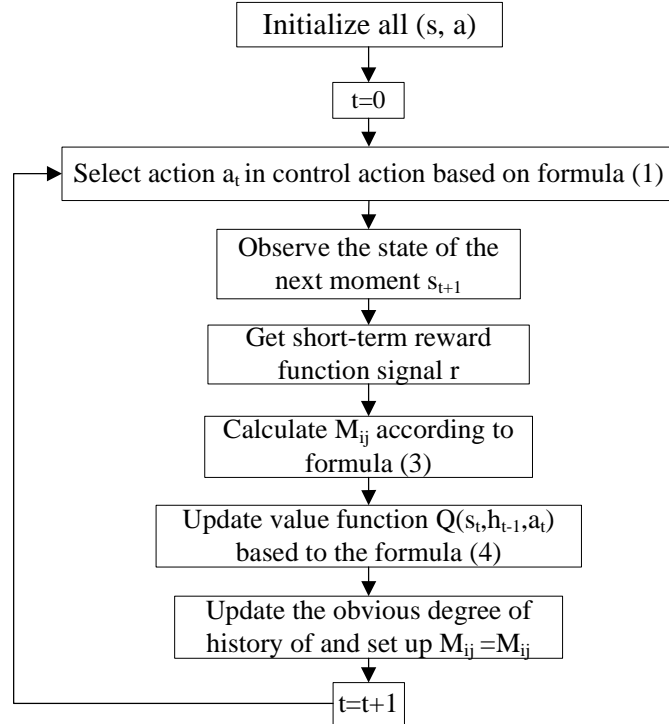


Figure 1. Iteration step of an improved multi-Agent Q learning algorithm

Simulation and Results

This emulator instance to IEEE 9-bus system as the basic framework^[8], as shown in Figure 2, where Node 1, Node 2 is the microturbine using fossil energy, node 9 using PV as power source. Simulation time is 10s, simulation step taken 0.05ms, we emulate the characteristic of the photovoltaic cell under the strongest light, which is the natural sunlight between 11a.m and 2 p.m.

We suppose the light intensity increases at 3s, and increase from $800\text{W} / \text{m}^2$ to $1000\text{W} / \text{m}^2$, at 6s Back to $800\text{W} / \text{m}^2$. The simulation results are shown in FIG. 3 - FIG. 6.

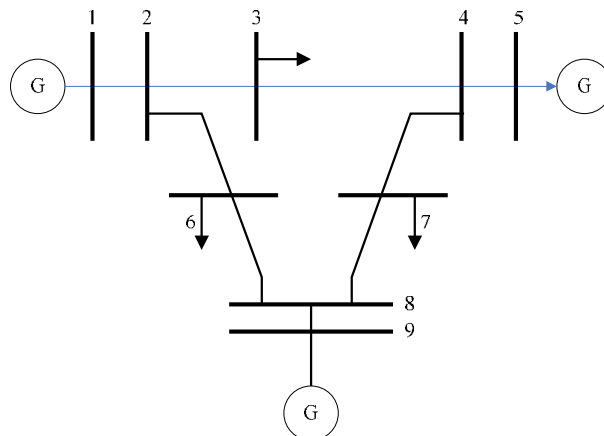


Figure 2 .EEE 9-bus system.

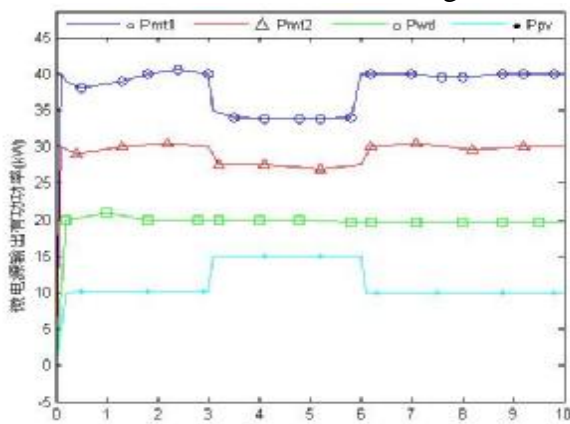


Figure 3. Active power output

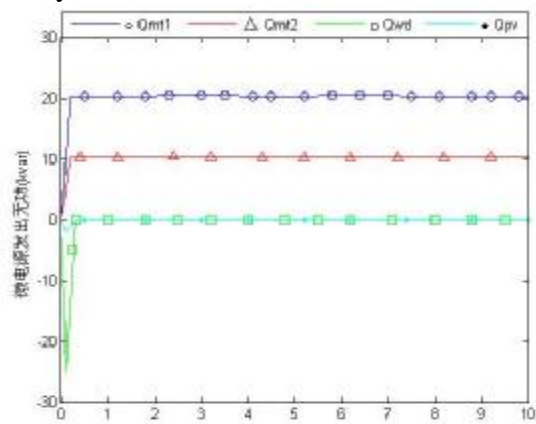


Figure 4. Reactive power output

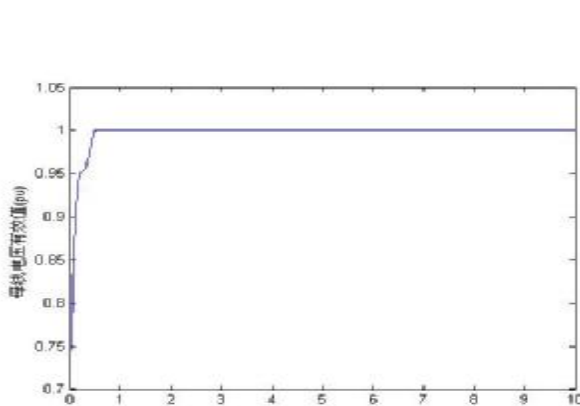


Figure 5. Bus voltage rms

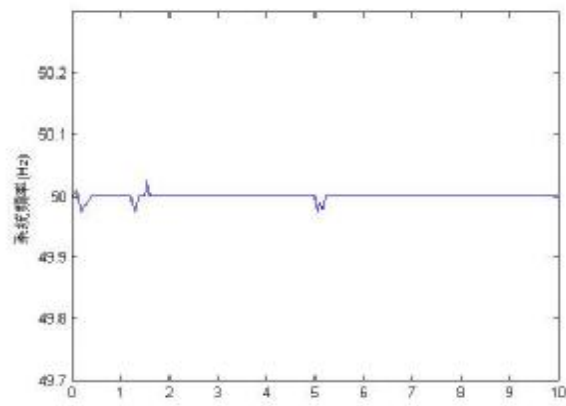


Figure 6. System frequency

The active power that is output at node1, after the initial slight fluctuations, is stabilized around 40kW, the reactive power at node1 is 25 VAR. Node 2 output active power, after the initial fluctuations, the power stays in 30kW, and the reactive power stays at 15Kvar. PV cell P/Q control shows excellent performance. Under light intensity of $800\text{W} / \text{m}^2$ the active power output is 11kW. When the light intensity changes, the output power increases with the increasing of the light intensity to 16kW. The reactive power changes after the light intensity slightly changes, and stabilized itself at 0k VAR. Meanwhile the active power output in node1 and node2 begins to decrease. In 6s due to the light intensity is reduced back to the original $800\text{W}/\text{m}^2$, node 1 and node2 begin to increase and stabilize themselves around 40kW and 30kW.

The simulation study shows that the agent can accurately manage the power of micro-power to provide reliable protection to the micro-grid. Simulation results show that, the use of multi-Agent Q learning algorithm can quickly restore the power of micro-grid to a stable state, when the power of the micro-grid fluctuates swiftly.

Conclusion

Based on Multi-Agent structure of micro-grid control system, this paper applies enhanced Multi-agent learning algorithm to the action-control process of each Agent. Then in order to solve Some problems of the traditional multi-Agent Q learning algorithm, this paper put forward a multi-Agent Q learning algorithm based on probability, and add the historical information as parameter to the process of selecting of optimal actions. Finally with the proposed algorithm this paper runs a simulation experiments in micro-grid system of IEEE9 .The results of the experiment show that this algorithm can restore the power of the micro-grid rapidly, when fluctuations occurs.

References

- [1] DUAN Tao, CHEN Wei-rong, et al. Reactive power optimization in power system based on multi-agent seeker optimization algorithm[J]. *Power System Protection and Control*, 2009, 37(14): 10-15.
- [2] A.L. Dimeas, N.D. Hatziargyriou. Agent based Control for Microgrids[C]. *Power Engineering Society General Meeting. IEEE 2007*:1-5
- [3] A.K.Basu,T.K.Panigrahi, S.Chowdhury, etal. Key energy management issues of setting market clearing price(MCP) in micro-grid scenario[C]. *42nd International Universities Power Engineering Conference . Sep. 4-6, 2007, Brighton, United Kingdom*:854—860.
- [4] Dimeas A L, Hatziargyriou N D. Operation of a multi-agent system for microgrid control[J]. *IEEE Transactions on Power Systems . 2005, 23(12)*: 50-57.
- [5] Katiraei F, Iravani M R. Power management strategies for a microgrid Systems with multiple distributed generation units[J]. *IEEE Transactions on Power*, 2006, 21(4) 1821-1831.
- [6] Guerrero J M, Matas J, De Vicuna L q et al. Wireless-control strategy for parallel operation of distributed-generation inverters[J]. *IEEE Transactions on Industry Electronics*, 2006, 53(5): 1461-1470.
- [7] Guerrero J M, Matas J, De Vicuna L G, et al. Decentralized control for parallel operation of distributed generation inverters using resistive output impedance[J]. *IEEE Transactions on Industry Electronics*, 2007, 54(2): 994-1004.
- [8] A. Chaouachi, R. M. Kamel, R. Andoulsi, and K. Nagasaka. Multiobjective intelligent energy management for a microgrid [J]. *IEEE Transactions on Industrial Electronics*, 2013, 60 (04): 1688-1699.