Multi-Agent Q(λ) Learning for Optimal Operation Management of Energy Internet

International Conference on Neural Information Processing

ICONIP 2017: Neural Information Processing pp 298-306 | Cite as

- Lingxiao Yang (1) Email author (ylxiao66@163.com)View author's OrcID profile (View OrcID profile)
- Qiuye Sun (1) View author's OrcID profile (View OrcID profile)
- Yue Han (2) View author's OrcID profile (View OrcID profile)

1. Northeastern University, Shenyang, China
2. State Grid Liaoning Electric Power Research Institute, Shenyang, China

Conference paper
First Online: 26 October 2017

Part of the Lecture Notes in Computer Science book series (LNCS, volume 10639)

Abstract

This paper proposes an optimal operation management methodology based on the multi-agent reinforcement learning (MARL) in energy internet (EI). An integrated approach to minimize the total cost of operation of such an electrical, natural gas and district heating network simultaneously is studied. A novel multi-agent Q(λ) learning method is presented to form a coordinated optimal management strategy of energy internet with multiple We-Energy(WE), and an equal interval sampling method is proposed to find the optimal discrete action sets so as to enhance the performance of the control areas. Furthermore, a global Q operator is designed to produce a global Q function considering the local reward from each agent which optimizes simultaneously. The proposed method verified by case studies applied to the modified energy network. Compared with the centralized approach, the test results show that the proposed method can provide a fast solution for the optimal operation management which can be applied to multiple We-Energy internet with sufficient accuracy.

Keywords

Multi-Agent reinforcement learning Optimal operation management Q(λ) learning Energy internet We-Energy

1 Introduction

The optimal operation management of energy has become an important socioeconomic problem in which exploring renewable energy and improving the supply efficiency are among important issues for realizing the EI of economic and security operation to reduce the network loss [1]. Under the development of renewable energy systems, the energy source will no longer be provided only in the form of the traditional energy supplier, and there will be several types of energy producers which means that users can choose the energy suppliers according to their own needs. In this supply mode, a novel energy accessing mode for energy Internet called We-Energy is proposed that it will be not only energy producer but also energy consumer [2]. In this regard, the cooperation among multiple energy forms and economic opportunities for enhancing the supply efficiency in EI was overlooked.

In [3, 4, 5, 6], the optimal power flow of electricity and natural gas combined system is discussed. Paper [7] proposed the generalized heuristic algorithm to study the optimal power flow of multiple energy system. While with the increasing utilization of co-generation plants, there is a challenge to find the optimal strategy in such a way for this class of complex nonlinear multi objective optimization problems. The domestic and foreign scholars made in-depth study in this question for more efficient algorithm. In addition to improve the basic heuristic algorithm for optimal energy flow [8], distributed algorithm has become a research focus [9].

Recently, reinforcement learning (RL) algorithm as a kind of machine learning algorithms attracts people's attention. Some learning strategies on the basis of RL to solve deterministic optimal control problems in continuous state spaces can be found in some studies such as [10, 11, 12, 13, 14, 15]. The MARL which is a new branch of reinforcement learning algorithm has been developed rapidly in various areas including distributed control, robotic teams, collaborative decision support systems, and economics [16]. MARL is defined to be composed of multiple agents, the whole system will achieve the learning goals through each agent executing part of reinforcement mission independently. All performance of MARL exhibits its advantage on the collaborative multi-task problems, and the features that aim at strategic decision make MARL widely used [17].

In this paper, the optimal operation management in EI is investigated at the transmission level. We propose an optimal model to improve the performance of environment for EI. The proposed formulation for multi-agent systems can be solved by machine learning algorithms which could find the optimization strategy intelligently. We present a MARL for distributed multicarrier energy network, which computes a global optimal policy in cooperative subsystems on the basis of the implementation of independent optimization for subsystems. A policy is defined as a set of actions deriving from the reward function connecting the environment. With the preprocessing method of action sets, the performance of the system will be improved.

This paper is organized as follows. In Sect. 2, the mathematical model of optimal management problem in EI is represented to improve the energy economy while Sect. 3 discussed about the multi-agent reinforcement learning methodology to solve this problem. Section 4 presents the case studies with a typical EI consists of several types of WEs. Finally, Sect. 5 concludes the discussion.

2 Optimal Operation Management Model of EI

In EI, different structure of prosumers will lead to a variety of supply and demand situations. WE, as a novel energy sub-region in EI, can exchange electricity, gas and heat from modern communications, power electronic conversion and automation technology with other WEs. As shown in Fig. 1, the main body of WE will be the individual, company or community that consists of energy production, user load or storage devices such as distributed generation, energy storage, CCHP and so on. WEs coordinate with each other to guarantee multi energy to reliable transport. Each WE is connected to be considered as the generalized node of EI.

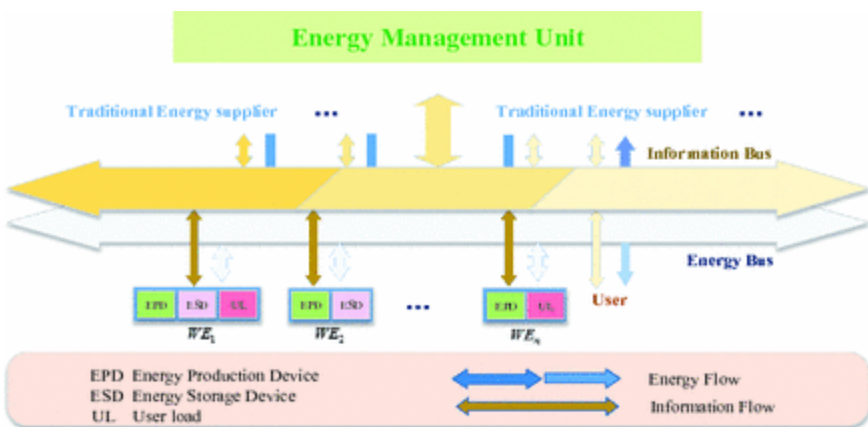

Fig. 1.

Energy flow model of WEs in energy internet.

In order to achieve the optimal energy flow of EI with multiple WEs, an operation management model is to find the optimal solution in a sense that required objective function is minimized while several equality and inequality constraints are satisfied. Through the coordinated optimization of WEs, the network will achieve the overall optimization.

The objective function in this paper is to optimize the operation cost which can be expressed by:

$$\min T = \sum_{i=1}^{n} F_i(P_i, F_\alpha)$$
$$s.t. \begin{cases} L = CP \\ G_\alpha(P_i) = 0 \\ \begin{cases} F_{\alpha\min} \le F_\alpha \le F_{\alpha\max} \\ P_{i\min} \le P_i \le P_{i\max} \end{cases} \end{cases}$$
(1)

Where $i = 1, 2, \ldots, n$, n is the number of WEs. P_i is the state variables of WE_i. F_α is the energy flow of multi-energy network including electric power network, natural gas network and centralized heating network. The first constraint stands for the input and output balance of an energy hub in WE which can be identified as a unit to achieve conversion and storage of multi-energy carriers. The second constraint and the third constrain respectively express the power balance equation of energy flow in EI and the variable limits to obtain a feasible operating point. And the last constrain is the limits of the network input. All the equality and inequality constraints which can refer to [18] will not be described in detail here.

For each WE, the total cost of operation in 24 h will be defined as the sum of the fuel consumed by WE which can be formulated as follows:

$$F_i = \sum_{t=1}^{24} \sum_{\beta=1}^{N} (a_\beta P_\beta(t) + b_\beta P_\beta^2(t))$$
(2)

Where t is the operation time. Generally, the cost of operation is formulated in a quadratic function on the basis of energy input power with several kinds of fuel such as natural gas, coal and so on. In addition, a_β, b_β are the unique coefficients for each fuel.

3 Proposed Methodology

Considering the energy network structure as well as the operation management model, the multi-agent reinforcement learning algorithm was applied to solve the problem innovatively.

3.1 Multi-Agent Q(λ) Learning Algorithm

MARL is a method which expands the single-agent RL. Each agent can obtain the rewards from adjacent agent with a few information. The global system use iteration to influence non-adjacent agent so as to optimize the performance of the whole system based on reinforcement learning.

According to the distributed structure of EI, each WE coordinates and interacts with each other to solve the complex problems. A optimal operation management of EI is a tuple $\langle n, \{S_i\}_{i=1,\ldots,n}, \{A_i\}_{i=1,\ldots,n}, \{R_i\}_{i=1,\ldots,n}, T\rangle$ where n is the number of WEs, S_i is the state space of WE_i, A_i is the action set of WE_i and $\{A_i\}_{i=1,\ldots,n}$ are the sets of actions available to the WEs, yielding the joint action set $A = A_1 + A_2 + \cdots + A_n$ that every WEs parallel compute for reinforcement learning. $T : S \times A \times S \to [0, 1]$ means the state transition probability. $R_i : S \times A \to R$ is the direct reward functions of each WE.

Given the MARL iteration rule for WE, the eligibility traces for the state-action will be updated as follows which can realize the delay control of signal during the optimization:

$$e_t(s,a) = \begin{cases} \gamma\lambda e_{t-1}(s,a) + 1 & if\ s = s_t\ and\ a = a_t \\ \gamma\lambda e_{t-1}(s,a) & otherwise \end{cases}$$

(3)

Where γ is the discount factor and λ is the recession coefficient. By adding the eligibility traces in the iteration, the problem of time reliability assignment in the optimization process will be solved.

The Q(λ) learning operation is defined as:

$$\delta_t = r_t(s, s', a_t) + \gamma \max Q_t(s', a_{t+1}) - Q_t(s, a_t)$$

(4)

$$Q_{t+1}(s,a) = Q_{t-1}(s,a) + \alpha\delta_t e_t(s,a)$$

(5)

Where α is the learning factor with $0 < \alpha < 1$, α indicates the proportion of update part in Q value.

3.2 Equal Interval Sampling for Action Sets

As a matter of fact, by considering the character of Markov Decision Processes for RL, the operational state of EI should be transformed into a discrete form. To do so, an equal interval sampling approach is given to each action set in such way that individuals with better fitness value improve the accuracy and generalization of computing.

In the MARL optimal operation management model, the energy hub is the key link of multiple energy coupling in each WE. Thus, for each control area, control variables studied in this paper is the power exchanged between the WEs and the hybrid energy network which contains electrical power P_e, natural gas power P_g and so on. Equal interval sampling approach for action sets is expressed as follows.

For P_e and P_g, the continuous attribute $a \in C$ and the range of P_a is $\min(P_a)$, $\max(P_a)$, the interval value can be designed as:

$$d = \frac{\max(P_a) - \min(P_a)}{k_a + 1}$$
(6)

Where the integer parameter k_a is the number of pre-set split points. And the split point set can be described as:

$$C_a^{k_a} = \begin{cases} \varnothing, & k_a = 0 \\ \{\min(V_a) + id, i = 1, \ldots, k_a\}, & k_a \geq 1 \end{cases}$$
(7)

If the integer parameter K_a is the maximum split points, split point set space can be expresses as:

$$\Omega_a = \{C_a^{k_a} : \ 0 \leq k_a \leq K_a\}$$
(8)

Therefore, the discrete action sets will be obtained by adjusting the parameters to find the optimal set of points. By preprocessing the action set, the learning speed will be accelerated and the accuracy of optimization could be improved.

3.3 Design of Reward Function for MARL

For MARL, in order to achieve the goal of optimal operation management, rewards in reinforcement learning should be combined with the objective function and constraint conditions.

The local immediate reward value R of each WE need satisfy the constraint conditions of power flow calculation to ensure the validity of the calculation results for each subsystem. Each WE will obtain the optimal strategy by maximizing reward function values. The local reward for WE is defined as formula (9).

Every WE will check control variables through connected transmission lines to see whether they meet the corresponding boundary conditions. If all of constraints are satisfied, the local reward signal will be set to the negative objective function. Otherwise, it will be zero. The local rewards are applied to each WE to guide action strategy.

$$R_{i,0}^{K} = \begin{cases} 0, & \textit{if constra}\text{int}\, s \textit{ are violated} \\ \frac{1}{F_i^K(X)}, & \textit{otherwise} \end{cases}$$
(9)

The aim of optimal operation management is to seek a best strategy from the action space, so that the global reward is presented as an average value of summation of local rewards from each WE. According to the global information discovery algorithm, the global reward signal can be discovered by using signals of node agents.

$$R^K = \frac{1}{n}\sum_{i=1}^{n} R_{i,0}^K$$

(10)

3.4 Implementation of the Algorithm for the Optimization of EI

The structure of MARL is shown below as Fig. 2. Through the multi-energy flow calculation for EI, the running status of each WE will be acquired. The following steps should be accomplished to solve the optimal energy flow problem in an EI using MARL algorithm.

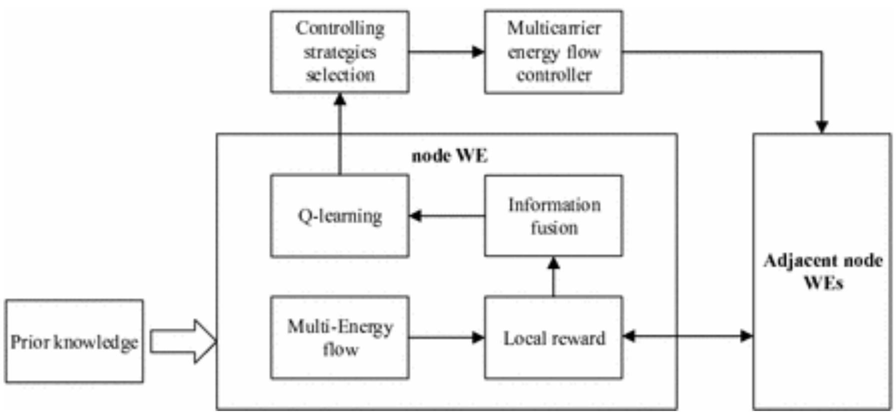

Fig. 2.

MARL optimization structure.

Each WE uses multi-energy flow calculation module to obtain the operation state. The local reward of WE will be obtained from the information interaction with environment according to the formula (9). Then, the global reward will be updated with the local reward if all the information is available in information fusion unit. The Q-learning unit will operate based on RL iteration rule to find the optimal strategy. Meanwhile, combined with the prior knowledge for initial action set, the learning state and learning efficiency could be improved. Each optimization iteration process contains four operation phases which includes detection of local signals, calculation of local reward function, exploration of the global signals, update the local Q matrix.

4 Simulations and Results

4.1 Test System

In this section, we apply the proposed multi-agent Q(λ) learning algorithm to the optimal energy flow problem for EI with the structure of 9-node internet with three WEs which is shown in Fig. 3. Percentages of the electric and heat demand in a 24 period compared to their basic values are illustrated in Fig. 4.

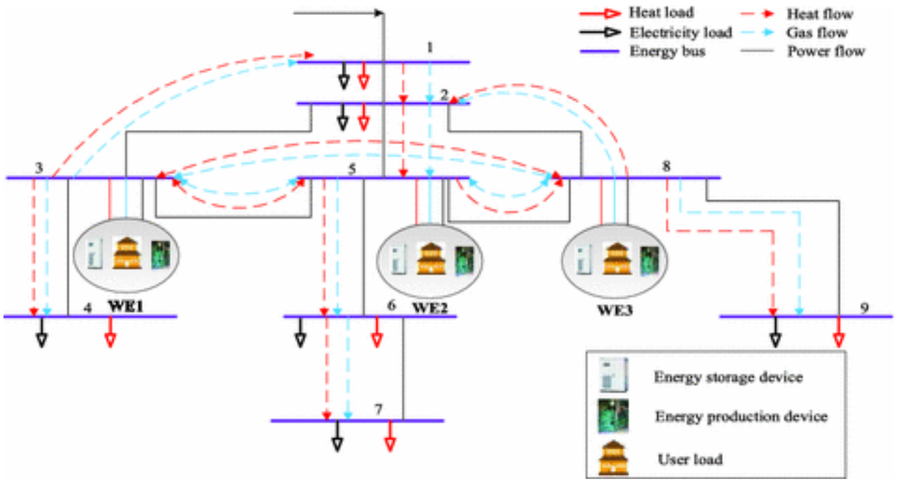

Fig. 3.

9-Node energy internet with three WEs.

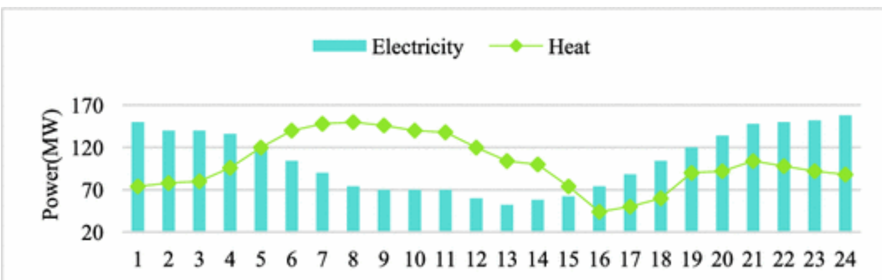

Fig. 4.

Electric and heat demand of EI during 24 h.

The interconnection between WEs includes information communication and energy transmission according to the demand. The energy coupling part is also achieved through the energy hub and its energy equipment from WEs includes energy storage device, energy production device and user load. The number of actions for MARL optimal power flow algorithm in this modified network is $120 + 136 + 114 = 370$ which is the sum of the action number from each WE according to the limits of input.

4.2 Simulation Results

As seen in Fig. 5, in order to achieve the overall energy costs and satisfy the needs of users, the optimal strategy is to obtain more energy from the power system, while reducing the natural gas from the natural gas system. And Fig. 6 shows the MARL convergence results at 1358.

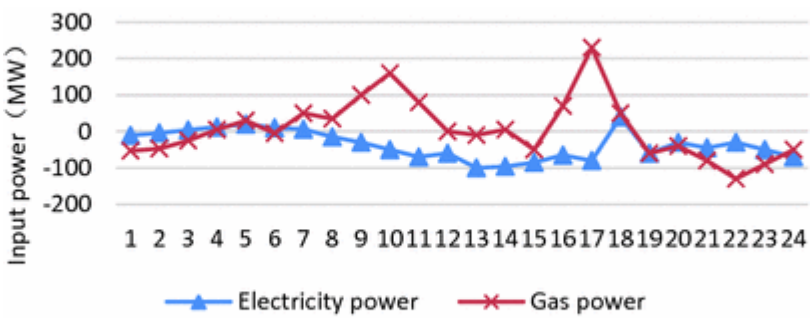

Fig. 5.

Input power of WE2 during 24 h.

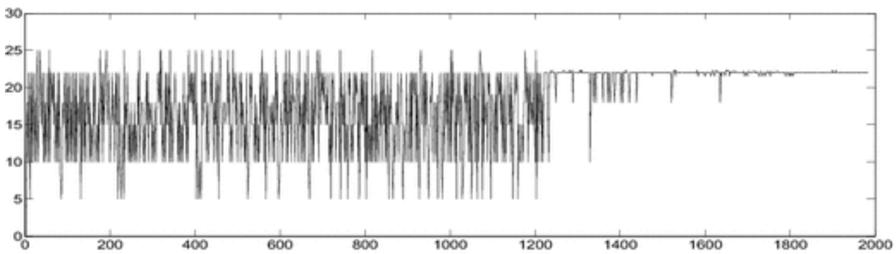

Fig. 6.

MARL process at $\alpha = 0.99, \gamma = 0.2$

Through the reinforcement learning of each WE, it can be concluded that the energy consumption of the EI is reduced by 27% under the constraints of the system.

5 Conclusion

This paper proposes an optimal operation management model and presents the multi-agent Q(λ) learning algorithm that can drive these agents to parallel learn behaviors. Compared with the centralized approach, they tend to reduce the calculation difficulty and require consideration of multiple aspects. Multi-Agent Reinforcement Learning is an effective way to improve the learning efficiency and solve the problem of “dimension disaster”. The method of utilizing each WE to undertake the task independently to reach the coordinated system is suitable for hierarchical control mode of energy internet. Above all, experimental results revealed that distributed algorithm has a faster convergence speed for the optimal energy flow model.

References

1. Shabanpour-Haghighi, A., Seifi, A.: Energy flow optimization in multicarrier systems. IEEE Trans. Ind. Inf. **11**(5), 1067–1077 (2016)
CrossRef (https://doi.org/10.1109/TII.2015.2462316)
Google Scholar (http://scholar.google.com
/scholar_lookup?title=Energy%20flow%20optimization%20in%20multicarrier%

20systems&author=A.%20Shabanpour-Haghighi&author=A.%20Seifi&journal=IEEE%20Trans.%20Ind.%20Inf.&volume=11&issue=5&pages=1067-1077&publication_year=2016)

2. Hu, J., Sun, Q., Teng, F.: A game-theoretic pricing model for energy internet in day-ahead trading market considering distributed generations uncertainty. In: International Conference on Computational Intelligence (SSCI), pp. 1–7. (2016)
Google Scholar (https://scholar.google.com/scholar?q=Hu%2C%20J.%2C%20Sun%2C%20Q.%2C%20Teng%2C%20F.%3A%20A%20game-theoretic%20pricing%20model%20for%20energy%20internet%20in%20day-ahead%20trading%20market%20considering%20distributed%20generations%20uncertainty.%20In%3A%20International%20Conference%20on%20Computational%20Intelligence%20%28SSCI%29%2C%20pp.%201%E2%80%937.%20%282016%29)

3. Correa-Posada, C.M., Sánchez-Martín, P.: Integrated power and natural gas model for energy adequacy in short-term operation. IEEE Trans. Power Syst. **30**(6), 3347–3355 (2015)
CrossRef (https://doi.org/10.1109/TPWRS.2014.2372013)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Integrated%20power%20and%20natural%20gas%20model%20for%20energy%20adequacy%20in%20short-term%20operation&author=CM.%20Correa-Posada&author=P.%20S%C3%A1nchez-Mart%C3%ADn&journal=IEEE%20Trans.%20Power%20Syst.&volume=30&issue=6&pages=3347-3355&publication_year=2015)

4. Alabdulwahab, A., Abusorrah, A., Zhang, X.: Coordination of interdependent natural gas and electricity infrastructures for firming the variability of wind energy in stochastic day-ahead scheduling. IEEE Trans. Sustain. Energy **6**(2), 606–615 (2015)
CrossRef (https://doi.org/10.1109/TSTE.2015.2399855)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Coordination%20of%20interdependent%20natural%20gas%20and%20electricity%20infrastructures%20for%20firming%20the%20variability%20of%20wind%20energy%20in%20stochastic%20day-ahead%20scheduling&author=A.%20Alabdulwahab&author=A.%20Abusorrah&author=X.%20Zhang&journal=IEEE%20Trans.%20Sustain.%20Energy&volume=6&issue=2&pages=606-615&publication_year=2015)

5. Xu, X., Jia, H., Chiang, H.D.: Dynamic modeling and interaction of hybrid natural gas and electricity supply system in microgrid. IEEE Trans. Power Syst. **30**(3), 1212–1221 (2015)
CrossRef (https://doi.org/10.1109/TPWRS.2014.2343021)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Dynamic%20modeling%20and%20interaction%20of%20hybrid%20natural%20gas%20and%20electricity%20supply%20system%20in%20microgrid&author=X.%20Xu&author=H.%20Jia&author=HD.%20Chiang&journal=IEEE%20Trans.%20Power%20Syst.&volume=30&issue=3&pages=1212-1221&publication_year=2015)

6. Zhang, X., Shahidehpour, M., Alabdulwahab, A.: Optimal expansion planning of

energy hub with multiple energy infrastructures. IEEE Trans. Smart Grid **6**(5), 2302–2311 (2015)
CrossRef (https://doi.org/10.1109/TSG.2015.2390640)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Optimal%20expansion%20planning%20of%20energy%20hub%20with%20multiple%20energy%20infrastructures&author=X.%20Zhang&author=M.%20Shahidehpour&author=A.%20Alabdulwahab&journal=IEEE%20Trans.%20Smart%20Grid&volume=6&issue=5&pages=2302-2311&publication_year=2015)

7. Shabanpour-Haghighi, A., Seifi, A.R.: Energy flow optimization in multicarrier systems. IEEE Trans. Industrial Inf. **11**(5), 1067–1077 (2015)
CrossRef (https://doi.org/10.1109/TII.2015.2462316)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Energy%20flow%20optimization%20in%20multicarrier%20systems&author=A.%20Shabanpour-Haghighi&author=AR.%20Seifi&journal=IEEE%20Trans.%20Industrial%20Inf.&volume=11&issue=5&pages=1067-1077&publication_year=2015)

8. Ríos-Mercado, R.Z., Borraz-Sánchez, C.: Optimization problems in natural gas transportation systems: a state-of-the-art review. Appl. Energy **147**, 536–555 (2015)
CrossRef (https://doi.org/10.1016/j.apenergy.2015.03.017)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Optimization%20problems%20in%20natural%20gas%20transportation%20systems%3A%20a%20state-of-the-art%20review&author=RZ.%20R%C3%ADos-Mercado&author=C.%20Borraz-S%C3%A1nchez&journal=Appl.%20Energy&volume=147&pages=536-555&publication_year=2015)

9. Sun, Q., Han, R., Zhang, H.: A Multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet. IEEE Trans. Smart Grid **6**(6), 3006–3019 (2015)
CrossRef (https://doi.org/10.1109/TSG.2015.2412779)
Google Scholar (http://scholar.google.com/scholar_lookup?title=A%20Multiagent-based%20consensus%20algorithm%20for%20distributed%20coordinated%20control%20of%20distributed%20generators%20in%20the%20energy%20internet&author=Q.%20Sun&author=R.%20Han&author=H.%20Zhang&journal=IEEE%20Trans.%20Smart%20Grid&volume=6&issue=6&pages=3006-3019&publication_year=2015)

10. Liu, D., Wei, Q.: Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. IEEE Trans. Neural Netw. Learn. Syst. **25**(3), 621–634 (2014)
CrossRef (https://doi.org/10.1109/TNNLS.2013.2281663)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Policy%20iteration%20adaptive%20dynamic%20programming%20algorithm%20for%20discrete-time%20nonlinear%20systems&author=D.%20Liu&author=Q.%20Wei&journal=IEEE%20Trans.%20Neural%20Netw.%20Learn.%20Syst.&volume=25&issue=3&pages=621-634&publication_year=2014)

11. Modares, H., Lewis, F.L., Naghibi-Sistani, M.B.: Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems. Automatica **50**(1), 193–202 (2014)
CrossRef (https://doi.org/10.1016/j.automatica.2013.09.043)
MATH (http://www.emis.de/MATH-item?1298.49042)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3157741)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Integral%20reinforcement%20learning%20and%20experience%20replay%20for%20adaptive%20optimal%20control%20of%20partially-unknown%20constrained-input%20continuous-time%20systems&author=H.%20Modares&author=FL.%20Lewis&author=MB.%20Naghibi-Sistani&journal=Automatica&volume=50&issue=1&pages=193-202&publication_year=2014)

12. Zhang, H., Cui, L., Zhang, X.: Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. IEEE Trans. Neural Netw. **22**(12), 2226–2236 (2011)
CrossRef (https://doi.org/10.1109/TNN.2011.2168538)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Data-driven%20robust%20approximate%20optimal%20tracking%20control%20for%20unknown%20general%20nonlinear%20systems%20using%20adaptive%20dynamic%20programming%20method&author=H.%20Zhang&author=L.%20Cui&author=X.%20Zhang&journal=IEEE%20Trans.%20Neural%20Netw.&volume=22&issue=12&pages=2226-2236&publication_year=2011)

13. Kamalapurkar, R., Dinh, H., Bhasin, S.: Approximate optimal trajectory tracking for continuous-time nonlinear systems. Automatica **51**, 40–48 (2015)
CrossRef (https://doi.org/10.1016/j.automatica.2014.10.103)
MATH (http://www.emis.de/MATH-item?1309.93086)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3284751)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Approximate%20optimal%20trajectory%20tracking%20for%20continuous-time%20nonlinear%20systems&author=R.%20Kamalapurkar&author=H.%20Dinh&author=S.%20Bhasin&journal=Automatica&volume=51&pages=40-48&publication_year=2015)

14. Wei, Q., Liu, D.: Adaptive dynamic programming for optimal tracking control of unknown nonlinear systems with application to coal gasification. IEEE Trans. Autom. Sci. Eng. **11**(4), 1020–1036 (2014)
CrossRef (https://doi.org/10.1109/TASE.2013.2284545)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Adaptive%20dynamic%20programming%20for%20optimal%20tracking%20control%20of%20unknown%20nonlinear%20systems%20with%20application%20to%20coal%20gasification&author=Q.%20Wei&author=D.%20Liu&journal=IEEE%20Trans.%20Autom.%20Sci.%20Eng.&volume=11&issue=4&pages=1020-1036&publication_year=2014)

15. Kamalapurkar, R., Andrews, L., Walters, P.: Model-based reinforcement learning for infinite-horizon approximate optimal tracking. IEEE Trans. Neural Netw. Learn. Syst. **28**(3), 753–758 (2017)
CrossRef (https://doi.org/10.1109/TNNLS.2015.2511658)

Google Scholar (http://scholar.google.com/scholar_lookup?title=Model-based%20reinforcement%20learning%20for%20infinite-horizon%20approximate%20optimal%20tracking&author=R.%20Kamalapurkar&author=L.%20Andrews&author=P.%20Walters&journal=IEEE%20Trans.%20Neural%20Netw.%20Learn.%20Syst.&volume=28&issue=3&pages=753-758&publication_year=2017)

16. Yu, C., Zhang, M., Ren, F.: Emotional multiagent reinforcement learning in spatial social dilemmas. IEEE Trans. Neural Netw. Learn. Syst. **26**(12), 3083–3096 (2015)
CrossRef (https://doi.org/10.1109/TNNLS.2015.2403394)
MathSciNet (http://www.ams.org/mathscinet-getitem?mr=3453259)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Emotional%20multiagent%20reinforcement%20learning%20in%20spatial%20social%20dilemmas&author=C.%20Yu&author=M.%20Zhang&author=F.%20Ren&journal=IEEE%20Trans.%20Neural%20Netw.%20Learn.%20Syst.&volume=26&issue=12&pages=3083-3096&publication_year=2015)

17. Plamondon, P., Chaib-draa, B., Benaskeur, A.R.: A Q-decomposition and bounded RTDP approach to resource allocation. In: Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems, p. 200. (2007)
Google Scholar (https://scholar.google.com/scholar?q=Plamondon%2C%20P.%2C%20Chaib-draa%2C%20B.%2C%20Benaskeur%2C%20A.R.%3A%20A%20Q-decomposition%20and%20bounded%20RTDP%20approach%20to%20resource%20allocation.%20In%3A%20Proceedings%20of%20the%206th%20International%20Joint%20Conference%20on%20Autonomous%20Agents%20and%20Multiagent%20Systems%2C%20p.%20200.%20%282007%29)

18. Shabanpour-Haghighi, A., Seifi, A.R.: Simultaneous integrated optimal energy flow of electricity, gas, and heat. Energy Convers. Manage. **101**, 579–591 (2015)
CrossRef (https://doi.org/10.1016/j.enconman.2015.06.002)
Google Scholar (http://scholar.google.com/scholar_lookup?title=Simultaneous%20integrated%20optimal%20energy%20flow%20of%20electricity%2C%20gas%2C%20and%20heat&author=A.%20Shabanpour-Haghighi&author=AR.%20Seifi&journal=Energy%20Convers.%20Manage.&volume=101&pages=579-591&publication_year=2015)

Copyright information

About this paper

Cite this paper as:

Yang L., Sun Q., Han Y. (2017) Multi-Agent Q(λ) Learning for Optimal Operation Management of Energy

Internet. In: Liu D., Xie S., Li Y., Zhao D., El-Alfy ES. (eds) Neural Information Processing. ICONIP 2017. Lecture Notes in Computer Science, vol 10639. Springer, Cham

- DOI (Digital Object Identifier) https://doi.org/10.1007/978-3-319-70136-3_32
- Publisher Name Springer, Cham
- Print ISBN 978-3-319-70135-6
- Online ISBN 978-3-319-70136-3
- eBook Packages Computer Science

- About this book
- Reprints and Permissions

Personalised recommendations