

# 考虑机组约束的时前市场发电商竞标策略

吴江, 李长山, 管晓宏, 高峰

(机械制造系统工程国家重点实验室(西安交通大学), 陕西省 西安市 710049)

## Unit Constraints Considered Genco's Bidding Strategies in Hour-ahead Electricity Market

WU Jiang, LI Chang-shan, GUAN Xiao-hong, GAO Feng

(State Key Laboratory for Manufacturing Systems Engineering(Xi'an Jiaotong University), Xi'an 710049, Shaanxi Province, China)

**ABSTRACT:** Power generation companies' (Genco) bidding strategy should consider the optimality in an entire period of time due to the unit operating constraints coupling over times. This paper presents a model for obtaining a Genco's optimal bidding strategy in the hour-ahead power market through  $Q$ -learning with unit operating constraints and start-up cost incorporated. The optimal bidding strategy in terms of cumulative total returns is gained through the iterative learning process. Numerical testing results show that this method is effective.

**KEY WORDS:** electricity market; bidding strategies; unit operating constraints;  $Q$ -learning

**摘要:** 由于机组约束的存在, 发电商在制定竞标策略时不能仅仅针对单个时间点进行优化, 还应考虑到这一决策对相邻时段的影响。该文在电能交易的时前市场背景下, 以  $Q$  学习算法为基础, 建立了考虑容量、爬升速率、最小开/关机时间等机组运行约束和机组启动成本的发电商竞标策略模型。该模型从市场前一小时的出清结果出发, 在考虑机组运行约束的前提下组织当前时刻机组的报价策略, 通过同环境的不断交互, 可以在随机波动较大的电力市场环境中得到当日累积回报最大的竞标决策。最后, 通过一个 10 机组系统对模型进行了仿真验证。

**关键词:** 电力市场; 竞标策略; 机组约束;  $Q$  学习

## 0 引言

在电力市场的竞争环境下, 电价主要由市场来调节。市场参与者通过提交自己的竞标数据进入市场参与竞争, 实现各自的运营目标。因此, 如何通

过优化报价策略使发电利润最大化, 成为了发电商关心的首要问题, 国内外学者在这方面也做了许多研究工作<sup>[1-6]</sup>。

由于存在着需求侧负荷波动、对手报价策略等随机因素的影响, 在实际的电力系统中, 市场环境很难进行精确建模, 发电商往往无法得到确定性的最优竞标决策。针对这一情况, 解决马尔可夫决策问题的增强学习(reinforcement learning)方法逐渐在策略竞标问题中得到应用。文献[7]利用增强学习中的 Roth-Erov 算法讨论了电力市场在复式拍卖(double auction)规则下的均衡价格; 文献[8-9]利用 Roth-Erov 算法比较了不同定价机制下的市场均衡价格; 通过将市场结算价格水平和发电商报价分别看作 agent 所处的状态和动作, 文献[10]建立了利用  $Q$  学习算法寻找发电商最优报价的策略模型; 文献[11]将  $Q(\lambda)$  学习算法引入发电商报价模型, 并以此对发电商的中长期报价策略进行了研究。

以上研究中, 只是针对单个时间点的竞标问题进行分析与优化, 而由于启停成本、启停时间、爬升约束等机组运行约束的存在, 发电商在制定竞标策略时, 不光要考虑当前时段的最优决策, 还要考虑这一策略对相邻时段的影响。因此, 在实际竞标问题中, 发电商应该加入对机组运行约束的考虑, 在一定的周期内得到最优的竞标策略。

本文以电能交易的时前市场为背景, 将增强学习中的  $Q$  学习算法引入到计及机组运行约束的发电商竞标策略问题, 考虑到了机组爬升约束、最小开/关机约束等运行约束和机组启动成本对竞标策略的影响, 建立了完整的  $Q$  学习策略模型。

## 1 问题描述

在本文中, 设市场为单一买主市场, 由电力调

基金项目: 国家自然科学基金项目(60274054); 国家 863 高技术基金项目(2004AA412050, 2007AA04Z154); 国家杰出青年科学基金(6970025)。

Project Supported by National Natural Science Foundation of China (60274054); The National High Technology Research and Development of China 863 Program (2004AA412050, 2007AA04Z154); Scientific Funds for Outstanding Young Scientists of China(6970025).

度中心(ISO)代表用户统一购买电能。市场交易按市场统一出清价的方式结算。电能市场按交易时间分为日前、时前、实时 3 个交易平台。在时前市场，发电商每小时申报一条针对下一时段的分段递增的阶梯状竞标曲线，每段由段容量和本段容量的申报段价组成。ISO 把各机组报价按段价从小到大排序，将满足下 1 小时市场预测负荷的最后一个机组的报价作为市场统一出清价(MCP)，所有报价在 MCP 以下的容量可以进行交易，购电方按 MCP 支付所购电量。

在  $t$  时刻，设下一时刻市场预测需求为  $D(t+1)$ ，机组  $i$  的最大出力为  $\bar{P}_i$ ，其  $M$  段竞标曲线  $\pi_i(t)$  可以用  $M$  组段容量一段价对  $\{\langle \alpha_{im}^t, \beta_{im}^t \rangle\}$  表示：

$$\pi_i(t) = \{\langle \alpha_{i1}^t, \beta_{i1}^t \rangle, \langle \alpha_{i2}^t, \beta_{i2}^t \rangle, \dots, \langle \alpha_{iM}^t, \beta_{iM}^t \rangle\} \quad (1)$$

式中：  $\sum_{m=1}^M \alpha_{im}^t \leq \bar{P}_i$ ；  $\beta_{i1}^t \leq \beta_{i2}^t \leq \dots \leq \beta_{iM}^t$ 。

机组  $i$  在所分配的电量  $p_i(t+1)$ ，市场的 MCP  $\lambda(t+1)$  分别为

$$\begin{aligned} &\langle \lambda(t+1), p_i(t+1) \rangle = G[\pi_i(t), \\ &\pi_2(t), \dots, \pi_i(t), \dots, \pi_l(t), D(t+1)] \quad (2) \end{aligned}$$

式中  $G(\cdot)$  即表示上文提到的 ISO 的市场价格出清机制和电量分配算法。

在市场环境下，发电商参与竞标的最终目的是发电利润最大化。当考虑机组的启停成本和运行约束时，机组  $i$  的竞标策略可以表述为如下优化问题：

$$\max \sum_{t=1}^T \{p_i(t)\lambda(t) - C_i[p_i(t)] - S_i[x(t), u(t)]\} \quad (3)$$

主要运行约束包括<sup>[12]</sup>：

(1) 状态转移约束。

$$\begin{cases} x_i(t+1) = x_i(t) + u_i(t), & x_i(t) \cdot u_i(t) > 0 \\ x_i(t+1) = u_i(t), & x_i(t) \cdot u_i(t) < 0 \end{cases} \quad (4)$$

(2) 容量约束。

$$\begin{cases} \underline{p}_i \leq p_i(t) \leq \bar{p}_i, & x_i(t) > 0 \\ p_i(t) = 0, & x_i(t) < 0 \end{cases} \quad (5)$$

(3) 爬升约束。

$$|p_i(t-1) - p_i(t)| \leq \Delta_i, \quad x_i(t) \geq 1 \text{ 且 } x_i(t+1) \geq 1 \quad (6)$$

(4) 最小开关机时间约束。

$$\begin{cases} u_i(t) = 1, & 1 \leq x_i(t) < \bar{\tau}_i \\ u_i(t) = -1, & -\bar{\tau}_i \leq x_i(t) < -1 \end{cases} \quad (7)$$

式(3)~(7)中： $C_i(p_i(t))$ 表示机组出力为 $p_i(t)$ 时的燃料成本； $p_i(t)$ 为机组在 $t$ 时段得到的最优竞标出力； $\bar{P}_i$ 、 $\underline{P}_i$ 为机组出力上/下限； $S_i(x_i(t), u_i(t))$ 为机组启动

成本； $T$ 为竞标周期； $u_i(t)$ 为机组在时段 $t$ 的离散决策变量，1表示开启状态，-1表示下关闭状态； $x_i(t)$ 为机组在时段 $t$ 的状态变量，表示机组已经开或关了几个时段； $\Delta_i$ 为机组爬升速率； $\bar{\tau}_i$ 为最小开机时间； $\underline{\tau}_i$ 为最小关机时间。

由式(2)可知，市场的MCP和机组所分配的电量不仅由自己的报价决定，市场需求和对手的策略同样会对结果产生较大影响。因此在市场环境下， $p_i(t)$ 与 $\lambda(t)$ 是无法事先精确给出的，故发电商竞标决策问题(2)~(7)属于随机优化问题，常规方法往往很难得到满意的结果。在本文中，利用非确定性环境下的Q学习方法来解决这一问题。

## 2 Q 学习策略模型

Q学习是一种研究马尔可夫决策过程的增强学习算法<sup>[13-15]</sup>。该算法不要求对市场环境进行精确建模，而是利用代表发电商的智能代理(agent)同市场环境不断的进行交互学习，得到累计回报达到最大的选择，从而使发电商得到最优的竞标策略。

Q学习算法中，将环境看作由若干个状态组成的集合 $S$ ，而每个状态 $s \in S$ 有它可以执行的动作集 $A_s$ 。在某一决策时间点 $t$ ，agent所处状态为 $s_t$ ，并在动作集中选取一个要执行的当前动作 $a_t$ 。环境响应此动作 $a_t$ ，给出回报 $r_t = r(s_t, a_t)$ ，并产生一个后继状态 $s_{t+1} = \delta(s_t, a_t)$ 。利用回报，agent通过定义Q函数来评估所选动作的优劣，即

$$Q(s, a) \equiv r(s, a) + \gamma \max_{a'} Q(\delta(s, a), a') \quad (8)$$

Q函数值为在状态 $s$ 执行动作 $a$ 的立即回报加上以后遵循最优策略的折算值。式中 $\gamma$ 为折算率。

在实际中，Q值通过agent同环境的不断交互学习获得更新，以逼近实际值

$$Q(s, a) \leftarrow r(s, a) + \gamma \max_{a'} Q(s', a') \quad (9)$$

从初始0时刻的状态 $s_0$ 出发到末时刻 $T$ 的状态 $s_T$ 为止的同环境的交互学习过程可看作一个情节(episode)。在一系列的情节中，通过对Q值的不断更新，agent只需考虑其当前状态 $s$ 下的每个可用动作 $a$ ，并选择其中使 $Q(s, a)$ 最大化的动作，就可以得到整个马尔可夫过程的最优决策。

在本文的竞标策略优化问题中，将负荷波动和对手策略看作发电商agent所需感知的非确定性环境，将机组 $i$ 作为竞标策略的研究对象。

假设爬升速率 $\Delta$ 可将 $i$ 的发电容量恰好平分成分成 $H$ 个点(本文以报价段 $M=3$ 为例，如果需要更精确

的输出, 可将分点作相应扩展), 加上停机状态, 机组的运行状态可以离散化为  $H+1$  个出力点。将  $t$  时刻机组的一个离散化出力点作为 agent 所处的一个状态  $s(t)$ , 即

$$s(t) = \begin{cases} s_0, & p_i(t) = 0 \\ s_h, & p_i(t) = \underline{P} + (h-1)\Delta \end{cases} \quad (10)$$

则  $t$  时刻的状态集合  $S(t) = \{s_0, s_1, \dots, s_H\}$ 。

由于受爬升约束的影响, 在离散化出力后, 在  $t$  时刻, 当状态  $s(t) = s_h$  给定时,  $t+1$  时刻机组的状态  $s(t+1)$  只可能是  $s_h$  或  $s_h$  的相邻状态  $s_{h-1}$  或  $s_{h+1}$ ; 当  $s_h$  为关机或最小开机状态时,  $s(t+1)$  的选择还要受到最小开/关机时间的约束。状态间的具体转移如图 1 所示。图中虚线表示转移只有当满足机组最小开/关机时间时才能实现。

机组  $i$  在决策周期  $T$  内的状态转移图可用图 2 表示。

当 agent 所处状态为  $s(t)$  时, 机组出力为  $p_i(t)$ 。将此时的所有可能的竞标决策作为  $s(t)$  状态所对应的动作集  $A_{s(t)}$ 。由图 1 可知, 受爬升约束的影响, 在出力的离散化模型中, 当  $t-1$  时刻的出力已知时, 机组在  $t$  时刻的出力最多只有 3 种可能。因此, 机

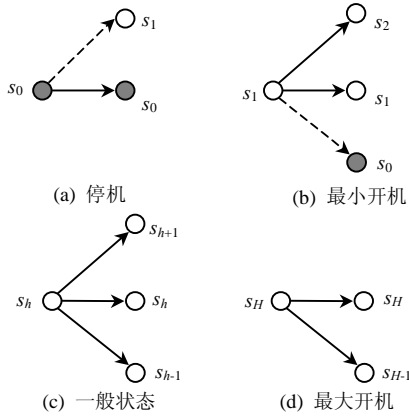


图 1 状态转移分类

Fig. 1 Four types of state transaction

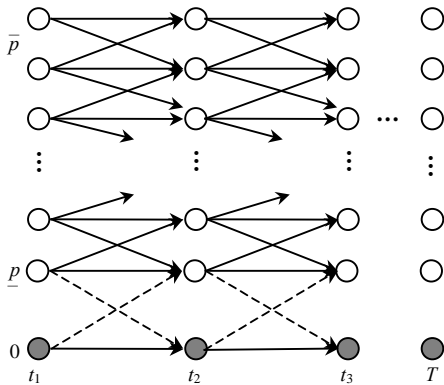


图 2 状态转移图

Fig. 2 State transaction diagram

组在  $t$  时刻所报竞标曲线的段容量也只可能有 3 段, 即竞标曲线分段数  $M=3$ 。在不同的状态可能的段容量如下:

(1)  $s(t) = s_0$ , 机组处于停机状态,  $p_i(t) = 0$ 。在满足最小关机时间时, agent 在下一时刻有开机  $s_1$  和继续停机  $s_0$  两种状态供选择; 当不满足最小关机时间时, agent 只能选择继续停机。因此在停机状态下, 竞标策略  $\pi_i(t)$  中的段容量  $\langle \alpha'_{i1}, \alpha'_{i2}, \alpha'_{i3} \rangle$  可以为  $\langle 0, \underline{P}, 0 \rangle$  或  $\langle 0, 0, 0 \rangle$ 。

(2)  $s(t) = s_1$ , 机组处于最小开机状态。在满足最小开机约束时, agent 在下一时刻可以选择  $s_0, s_1$  和  $s_2$  3 种状态, 而不满足最小开机时间时, agent 不能选择  $s_0$  状态。则可选的段容量为  $\langle 0, 0, 0 \rangle, \langle 0, \underline{P}, \Delta \rangle$  或  $\langle \underline{P}, \Delta, 0 \rangle$ 。

(3)  $s(t) = s_H$ , 机组处于最大开机状态。agent 在下一时刻可以选择  $s_H$  和  $s_{H-1}$  2 种状态, 则段容量为  $\langle \bar{P} - \Delta, \Delta, 0 \rangle$ 。

(4)  $s(t) = s_h$ , 机组处于一般运行状态。agent 在下一时刻可以选择  $\{s_{h-1}, s_h, s_{h+1}\}$  3 种状态, 则段容量为  $\langle p_i(t) - \Delta, \Delta, \Delta \rangle$ 。

为段容量  $\langle \alpha'_{i1}, \alpha'_{i2}, \alpha'_{i3} \rangle$  配以不同策略下的申报价格  $\langle \beta'_{i1}, \beta'_{i2}, \beta'_{i3} \rangle$  就构成了机组  $i$  在  $t$  时刻的竞标策略集  $\{\pi_{ie}(t)\}_E$ , 即 agent 在状态  $s(t)$  的动作集合  $A_{s(t)}$ 。

在  $A_{s(t)}$  中, 随机选取动作  $a_{s(t)} = \pi_{ie}(t)$  作为  $t$  时刻的报价申报给市场环境, 通过对其他发电商和 ISO 的市场出清机制的模拟, 环境响应动作  $a_{s(t)}$  并返回  $t+1$  时刻的 MCP  $\lambda(t+1)$  与分配发电量  $p_i(t+1)$ , 并由此得到机组  $i$  在  $t+1$  时刻的利润为

$$R_i(t+1) = \lambda(t+1)p_i(t+1) - C_i[p_i(t+1)] - S_i[x(t+1), u(t+1)] \quad (11)$$

将  $R_i(t+1)$  作为市场环境对 agent 的回报  $r[s(t), a(t)]$ 。根据式(12), 由  $p_i(t+1)$  可以得到  $t+1$  时刻的状态转移  $s(t+1)$ , 并产生新的竞标动作  $a_{s(t+1)} = \pi_{ie}[s(t+1)]$  申报给市场环境, 得到环境响应的回报  $r[s(t+1), a(t+1)]$ 。利用图 2 中的状态转移关系, 不断选择报价动作, 即可得到发电商在整个报价周期  $T$  内的报价。

不断重复以上情节, 通过 agent 同环境的不断交互学习, 就可以得到发电商在报价周期  $T$  内的最优竞标策略。

利用以上状态-转移模型, 上节中的发电商竞标策略问题(2)~(7)可转化为如下  $Q$  学习问题:

$$\pi^*[s(t)] = \arg \max_{a(t)} \{r[s(t), a(t)] + V^*[s(t+1)]\} \quad (12)$$

式中： $\pi^*[s(t)]$ 为 $s(t)$ 状态的最优竞标策略； $V^*[s(t+1)]$ 为 $t+1$ 时刻后的最大累积回报，即

$$V[s(t+1)] = \sum_{j=t+1}^T R_i(j) \quad (13)$$

则在本问题中， $Q$ 函数更新可表示为

$$Q_n[s(t), a(t)] \leftarrow (1 - \nu_n) Q_{n-1}[s(t), a(t)] + \nu_n \{R_i(t+1) + \max_{a(t+1)} Q_{n-1}[s(t+1), a(t+1)]\} \quad (14)$$

其中  $\nu_n = 1/[1 + \nu_n[s(t), a(t)]] \quad (15)$

式中 $\nu_n[s(t), a(t)]$ 为状态-动作对 $\langle s(t), a(t) \rangle$ 在 $n$ 次循环内被访问的次数。

在状态 $s(t)$ ，利用 $Q$ 值随机选择集合 $A_{s(t)}$ 中的竞标策略作为动作 $a(t)$ 。各个动作的选择概率由下式决定：

$$P(a_j(t) | s(t)) = \frac{k^{Q(s(t), a_j(t))}}{\sum_e k^{Q(s(t), a_e(t))}} \quad (16)$$

式中： $k$ 为动作选择因子，决定了选择优先考虑高 $Q$ 值的程度：较大的 $k$ 值会将较高的概率赋予当前 $Q$ 值大的动作，致使 agent 更倾向于选择利用它已经学习过的、会产生高汇报的动作，即采用利用型策略；而较小的 $k$ 值会使减少 $Q$ 值的高低所带来的选择概率的差别，致使 agent 探索那些未知的、当前 $Q$ 值还不高的动作，即采用探索型策略。在本文中，随迭代次数的变化逐渐增加 $k$ 值，以使 agent 在学习的早期采用探索型策略，然后逐步转换到利用型策略。

### 3 算例分析

本文采用一个 10 机组系统模拟市场环境进行仿真测试。机组参数及成本数据如表 1、表 2 所示，其中 1<sup>#</sup>、2<sup>#</sup>为具有爬升约束的机组。另外，以 24 h 为考虑机组运行约束的竞标周期。

由于电力市场基本属于寡头竞争市场，市场力的存在会对市场的交易结果造成较大的影响，因此在算例中考虑市场力强和市场力较弱的 2 种负荷需求状况。2 种系统负荷需求曲线见图 3。当负荷曲线为 $D_1$ 时，在高峰时段系统需求远远高于除去 1<sup>#</sup>机组容量后的系统剩余容量，大容量机组(如 1<sup>#</sup>、2<sup>#</sup>、3<sup>#</sup>)可通过报价滞留操纵市场价格；而当负荷曲线为 $D_2$ 时，系统需求维持在较低的水平，任何机组操纵市场的行为都会面临出力减少甚至停机的风险。

表 1 机组参数

Tab. 1 Unit's operating constraints

机组编号	$p$ /MW	$p$ /MW	$\Delta$ /MW	$\tau$ /h	$\tau$ /h	$\tau$ /h
1 <sup>#</sup>	150	450	60	8	8	13
2 <sup>#</sup>	150	470	40	8	8	13
3 <sup>#</sup>	50	180	—	5	5	9
4 <sup>#</sup>	50	180	—	5	5	9
5 <sup>#</sup>	100	350	—	6	6	10
6 <sup>#</sup>	40	120	—	3	3	6
7 <sup>#</sup>	25	85	—	3	3	5
8 <sup>#</sup>	10	55	—	1	1	1
9 <sup>#</sup>	10	55	—	1	1	1
10 <sup>#</sup>	10	55	—	1	1	1

表 2 机组成本

Tab. 2 Unit's operating cost

机组编号	燃料成本/\$			启动成本/\$	
	$a$	$b$	$c$	$S^b$	$S^c$
1 <sup>#</sup>	0.000 48	15.69	930	4 500	9 000
2 <sup>#</sup>	0.000 31	16.16	890	5 000	10 000
3 <sup>#</sup>	0.002 75	19.31	700	550	1 100
4 <sup>#</sup>	0.002 11	19.57	680	560	1 120
5 <sup>#</sup>	0.001 98	18.07	810	900	1 800
6 <sup>#</sup>	0.007 12	21.26	370	170	340
7 <sup>#</sup>	0.000 79	23.74	480	260	520
8 <sup>#</sup>	0.004 13	24.92	660	30	60
9 <sup>#</sup>	0.002 22	25.27	665	30	60
10 <sup>#</sup>	0.001 73	25.79	670	30	60

在仿真环境中，机组的竞标策略包括 $Q$ 学习策略、固定报价策略、随机报价策略 3 种。对于采用 $Q$ 学习策略的机组，将机组平均边际成本的 1 倍、1.2 倍、1.5 倍作为低(L)中(M)高(H)3 档报价同段容量进行组合，构成竞标策略动作集 $A_{s(t)}$ 。如对于一般的 3 段报价，段价集 $\{\langle d_{i1}, d_{i2}, d_{i3} \rangle\}$ 共有{LLL, LLM, LLH, LMM, LMH, LHH, MMM, MMH, MHH, HHH} 10 种选择。固定报价策略是在机组边际成本基础上申报递增的竞标曲线，随机报价策略是在固定报价的基础上加入随机扰动，固定报价策略和随机报价策略分别用于模拟下面算例中的确定性市场环境和非确定性市场环境。

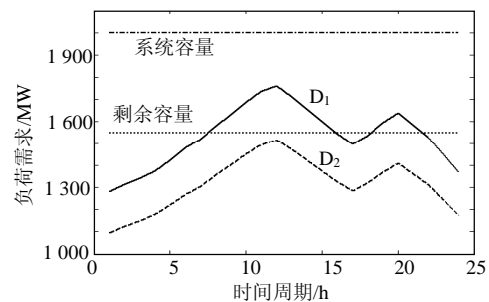


图 3 仿真环境的系统负荷需求  
Fig. 3 System demands in simulations

(1) 市场环境的随机性对 $Q$ 学习策略的影响。正如前文所提到的，系统负荷波动、对手的竞

标行为等非确定性因素使得电力市场交易具有较大的随机性,本算例将讨论不同程度的随机环境对 $Q$ 学习策略的影响。算例分为确定性、较小随机和较大随机3种仿真环境。在确定性环境中,系统负荷为 $D_1$ 已知,1#机组采用 $Q$ 学习策略,其他9个机组采用固定报价方式;在较小随机环境中,系统负荷在 $D_1$ 的基础上按5%的相对方差随机扰动,1#机组采用 $Q$ 学习策略,其他9个机组采用固定报价方式;在较大随机环境中,系统负荷在 $D_1$ 的基础上按10%的相对方差随机扰动,1#机组采用 $Q$ 学习策略,其他9个机组采用相对方差为5%的随机报价方式。1#机组的初始出力 $P_0=330$  MW。

图4由上到下分别表示在确定性环境、较小随机环境和较大随机环境下1#机组在整个竞标周期的总利润。由图中可以看出,采用 $Q$ 学习策略的1#机组利润在仿真初期波动较大,而随着仿真次数的增加, $Q$ 学习算法逐渐收敛,利润也趋于稳定,并在最后维持在较高的水平。当市场环境的随机性增大时,总利润的波动性也会变大,但 $Q$ 学习算法仍能收敛,得到较好的策略结果。

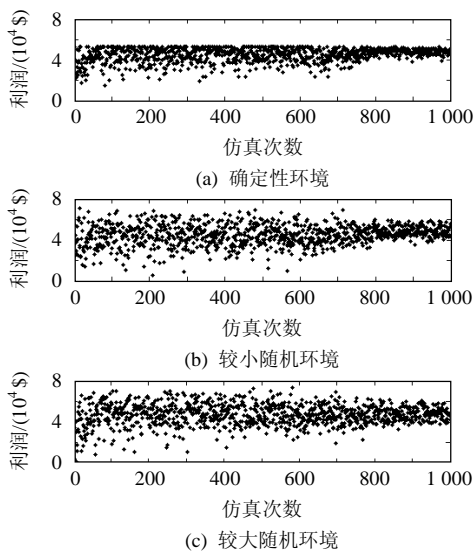


图4 不同市场环境下1#机组总利润

Fig. 4 Total rewards of 1# unit in different environments

(2) 负荷需求对 $Q$ 学习策略的影响。

不同的负荷需求水平将会导致不同的市场力表现,本算例针对图3中 $D_1$ 、 $D_2$ 两种负荷水平,讨论在不同的负荷需求下的 $Q$ 学习策略。算例中采用较小随机环境,即系统负荷按5%的相对方差随机扰动,1#机组采用 $Q$ 学习策略,其他9个机组采用固定报价方式。1#机组的初始出力 $P_0=330$  MW。1#机组60 MW的爬升约束将150~450 MW的出力区

间分为6个出力水平,加上停机状态,在 $Q$ 学习算法中1#机组共有7个状态可供选择,如图5中的细点所示。图5中圆圈表示在 $D_1$ 和 $D_2$ 两种系统需求下1#机组最终收敛的状态。当需求为 $D_1$ 时系统负荷趋紧,通过竞标策略的选择,1#机组的出力始终保持较高的水平,并且在中午的负荷高峰投入了全部的发电容量;在 $D_2$ 情况下平均需求仅有 $D_1$ 时的80%,但通过 $Q$ 学习算法优化竞标策略,仍然保证了在高峰时段1#机组的满负荷发电。而对于需求偏低的夜间时段,在两种情况下agent均选择了逐渐减少出力。由图6的利润比较也可以看出, $D_1$ 情况下24 h的总利润一直维持在较高的水平,并最终趋于收敛;在 $D_2$ 情况下,虽然由于需求较少导致竞争加剧,总利润相对 $D_1$ 情况降低,但随着迭代次数的增加,agent还是找到了增加回报的最优策略,使得总利润不断增加,最终收敛到了一个较高的水平。此算例的详细报价结果见表3、4。

表3和表4分别表示了第3节算例2 $D_1$ 和 $D_2$

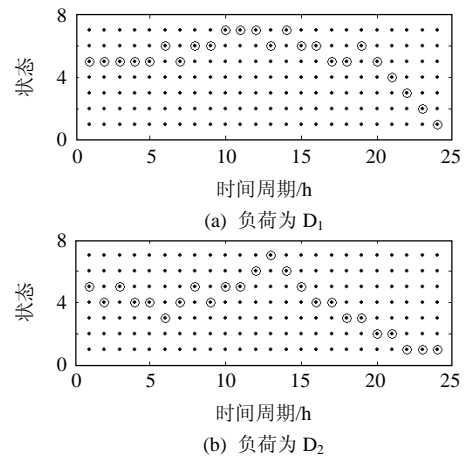


图5 不同系统需求下1#机组的收敛状态  
Fig. 5 Convergence state of 1# unit under different demands

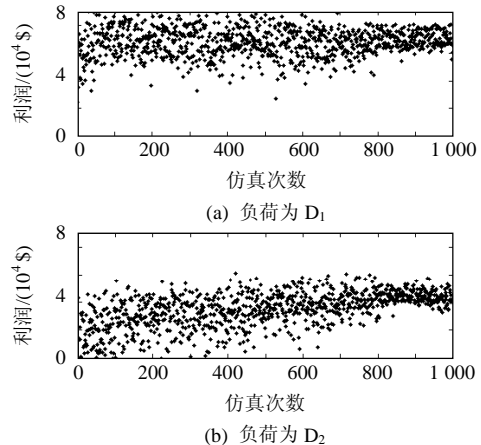


图6 不同系统需求下1#机组总利润  
Fig. 6 Total rewards of 1# unit under different demands

表 3 负荷为D<sub>1</sub>时 1<sup>#</sup>机组各状态最优报价决策  
**Tab. 3 Optimal bidding decisions of unit 1<sup>#</sup> in each state with demand D<sub>1</sub>**

时间/h	P/MW						
	0	150	210	270	330	390	450
1	—	—	—	—	<b>LLM</b>	—	—
2	—	—	—	LMM	MMM	<b>LLH</b>	—
3	—	—	LMH	LLL	<b>MMM</b>	LLH	LM
4	—	LM	LMH	LLL	LLH	<b>MMM</b>	MH
5	—	MM	LMH	LLM	LMM	MMM	<b>LM</b>
6	—	—	LMM	LMM	MHH	MMH	<b>LH</b>
7	—	MM	LLL	LHH	LMH	LHH	<b>LH</b>
8	—	LL	LLL	LLL	LHH	LHH	<b>LM</b>
9	—	—	MMM	LMM	LMH	<b>LLH</b>	MH
10	—	—	LMH	HHH	LHH	<b>LLH</b>	MH
11	—	—	—	MHH	LLM	LHH	<b>LM</b>
12	—	—	—	—	LLH	LMM	<b>HH</b>
13	—	—	—	—	LLM	HHH	<b>MH</b>
14	—	—	—	LMH	LMH	MHH	<b>MH</b>
15	—	—	—	LMM	LLL	LHH	<b>MH</b>
16	—	—	—	LMM	MHH	<b>MHH</b>	MH
17	—	—	LLM	LMM	MHH	<b>LMH</b>	LM
18	—	—	LLH	LHH	LLH	<b>MMH</b>	LH
19	—	LM	MMM	HHH	MHH	<b>LMH</b>	MH
20	—	—	MMH	MHH	LMH	<b>LHH</b>	MH
21	—	—	LLM	MHH	<b>LHH</b>	HHH	MM
22	—	—	LMH	<b>LHH</b>	LLL	LMH	MH
23	—	LH	<b>LHH</b>	LHH	MMM	LLM	MM
24	—	<b>MM</b>	LMM	LMM	LLL	MMM	LM

表 4 负荷为D<sub>2</sub>时 1<sup>#</sup>机组各状态最优报价决策  
**Tab. 4 Optimal bidding decisions of unit 1<sup>#</sup> in each state with demand D<sub>2</sub>**

时间/h	P/MW						
	0	150	210	270	330	390	450
1	—	—	—	—	<b>LLL</b>	—	—
2	—	—	—	<b>LLL</b>	LLL	LLM	—
3	—	—	LLM	<b>LLL</b>	LMH	LMH	LH
4	—	LL	LLM	<b>MMM</b>	LMH	LLH	—
5	0	MM	LMM	<b>LLL</b>	LLM	MMH	LH
6	0	LL	LLL	MMM	<b>LLL</b>	LLL	LM
7	0	LM	MMM	MMM	<b>LLL</b>	LLM	LH
8	0	LL	LLL	LLM	<b>MMH</b>	LLL	LM
9	0	MM	LMH	MMM	LLL	<b>LMH</b>	MH
10	0	LM	LLM	MMM	LHH	<b>MHH</b>	MH
11	0	LL	LMM	LLH	MHH	<b>MHH</b>	LH
12	0	LL	LMM	MMM	LMH	LHH	<b>MH</b>
13	M	LL	LMM	LLH	LHH	MHH	<b>MH</b>
14	M	LM	LMM	LMH	LMM	MHH	<b>MH</b>
15	H	LL	LMM	LMH	LLH	<b>MHH</b>	MM
16	0	LL	MMM	LLM	LHH	MMM	<b>MM</b>
17	H	LL	LMM	LHH	LLH	<b>MMM</b>	MH
18	0	MM	MMM	MMM	MMH	<b>MMM</b>	LL
19	H	LL	LMM	LHH	MHH	<b>LMH</b>	MH
20	0	LL	LMH	MMM	MMH	<b>MHH</b>	LH
21	H	MH	MMM	LLH	<b>LHH</b>	LHH	MM
22	0	MH	LHH	<b>LHH</b>	MMM	LHH	MM
23	0	LH	<b>MHH</b>	MHH	LMM	MMM	MH
24	0	<b>MM</b>	LMH	LLL	LLH	LLL	LL

两种负荷水平的市场环境中，在Q学习策略下，日调度周期内 1<sup>#</sup>机组在各个状态下的最优报价决策。其中，“L”、“M”、“H”代表了各段价的价格水平；“0”代表机组关机；“—”代表此状态未出现。表中加粗状态为图 5 中所示的最终收敛状态。

### 4 结论

本文根据市场统一出清价结算、发电商分段报价的电能交易时前市场规则，提出了基于 Q 学习算法的发电商优化竞标模型。该模型加入了对机组爬升约束、最小开/关机约束等运行约束和机组启动成本对竞标策略影响的考虑，得出了在一定的竞标周期内发电商的最优竞标决策。最后利用一个 10 机组系统为例对算法模型进行了讨论。从仿真结果看，agent 能够灵活的根据市场环境的反馈选择竞标策略，而且即便是在波动较大的随机环境下，该算法仍能得到较好的结果。

### 参考文献

- [1] David A K. Competitive bidding in electricity supply[J]. IEE Proceedings of Generation, Transmission and Distribution, 1993, 140(5): 421-426.
- [2] David A K, Fushuan Wen. Strategic bidding in competitive electricity markets: a literature survey[C]. IEEE Power Engineering Society Summer Meeting, Seattle, 2000.
- [3] Guan X, Ho Y C, Lai F. An ordinal optimize based bidding strategy for electric power suppliers in the daily energy market[J]. IEEE Transaction on Power Systems, 2001, 16(4): 788-797.
- [4] 高鑫, 王秀丽, 雷兵, 等. 独立发电商的报价策略研究[J]. 中国电机工程学报, 2004, 24(7): 41-46. Gao Xin, Wang Xiuli, Lei Bing, et al. Research on bidding strategy for an independent power plant[J]. Proceedings of the CSEE, 2004, 24(7): 41-46(in Chinese).
- [5] 李益国, 沈炯, 刘西陲. 基于机会约束规划的发电公司竞标策略[J]. 中国电机工程学报, 2006, 26(10): 120-123. Li Yiguo, Shen Jiong, Liu Xichui. A new bidding strategy for power plants based on chance-constrained programming[J]. Proceedings of the CSEE, 2006, 26(10): 120-123(in Chinese).
- [6] 吕泉, 武亚光, 李卫东. 基于多心理账户的发电商报价决策模型[J]. 中国电机工程学报, 2006, 26(24): 24-30. Lü Quan, Wu Yaguang, Li Weidong. A novel bid decision-making model for generation companies based on multiple mental account [J]. Proceedings of the CSEE, 2006, 26(24): 24-30 (in Chinese).
- [7] Nicolaisen J, Petrov V, Tesfatsion L. Market power and efficiency in a computational electricity market with discriminatory double-auction pricing[J]. IEEE Transactions on Evolutionary Computation, 2001, 5(5): 504-523.
- [8] 邹斌, 李庆华, 言茂松. 电力拍卖市场的智能代理仿真模型[J]. 中国电机工程学报, 2005, 25(15): 7-11. Zou Bin, Li Qinghua, Yan Maosong. An agent-based simulation model on pool-based electricity market using locational marginal

- price[J]. Proceedings of the CSEE, 2005, 25(15): 7-11(in Chinese).
- [9] 邹斌, 言茂松, 谢光前. 不同定价方式下电力拍卖市场运行特征的比较研究[J]. 电力系统自动化, 2005, 28(15): 7-14.  
Zou Bin, Yan Maosong, Xie Guangqian. Comparisons among pricing methods in pool-based electricity market by agent-based simulation [J]. Automation of Electric Power System, 2005, 28(15): 7-14(in Chinese).
- [10] Xiong G F, Hashiyama T, Okuma S. An electricity supplier bidding strategy through  $Q$ -learning[C]. IEEE Power Engineering Society Summer Meeting, Chicago, 2002.
- [11] 马豫超, 蒋传文, 候志俭, 等. 基于自加强学习算法的发电商报价策略研究[J]. 中国电机工程学报, 2006, 26(17): 12-17.  
Ma Yuchao, Jiang Chuanwen, Hou Zhijian, et al. Strategic bidding of the electricity producers based on the reinforcement learning [J]. Proceedings of the CSEE, 2006, 26(17): 12-17(in Chinese).
- [12] Guan X, Luh P B, Yan H, et al. An optimization based method for unit commitment[J]. International Journal of Electrical Power & Energy Systems, 1992, 14(1): 9-17.
- [13] Watkins C J C H. Learning from delayed rewards[D]. Cambridge: Cambridge University, 1989.
- [14] Watkins C J C H, Dayan P.  $Q$ -learning[J]. Machine Learning, 1992, 8(3-4): 279-292.
- [15] Gosavi A. Reinforcement learning for long-run average cost[J]. European Journal of Operational Research, 2004, 155(3): 654-674.



吴江

收稿日期: 2007-09-10。

作者简介:

吴江(1980—), 男, 博士研究生, 主要从事电力市场方面的研究, jwu@sei.xjtu.edu.cn;

李长山(1982—), 男, 硕士研究生, 主要从事电力市场仿真方面的研究;

管晓宏(1955—), 男, 教授, 博士生导师, 主要从事电力市场、电力与制造系统优化调度、计算机网络信息安全等方面的研究;

高峰(1967—), 男, 教授, 主要从事电力市场预测, 优化与竞标等方面的研究工作。

(编辑 王彦骏)