Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Designing a Bidding-Agent for Electricity Markets: A Multi Agent Cooperative Learning Approach

Ali Nouri Dariani*, Arastoo Fazeli Neishabour, Ashkan Rahimi-Kian*, Maziar Ahmad Sharbafi***

** CIPCE, School of ECE, College of Eng., University of Tehran, Iran*
emails: a.nouri@ece.ut.ac.ir, arkian@ut.ac.it, m.sharbafi@ut.ac.ir
***School of Electrical Engineering, Sharif University of Technology, Tehran, Iran*
email: arastoo_fazeli@ee.sharif.edu

Abstract: In the restructuring process of power systems, bidding strategies are the main routes for making more profit and therefore, there has been a wide research on them. In this paper, we consider a bidding model that is based on the residual demand (RD). Our approach concerns the identification of RD and learning how to bid according to it. When the agents bid in a market, some knowledge about the environment is obtained, which may be more influential than the obtained sheer profit. So, the agent should learn when and how to pay attention to the environment. The agent's expertise is measured in each part of the residual demand curve and, then its shortcomings are identified. Our agent makes a balance between the profit making and knowledge increasing processes. The designed agent's mind consists of many subagents, each learning its own task and also cooperating with others simultaneously. In this regard, a credit assignment system is implemented among the agents and the cooperative learning trends are applied. Finally, through a few case studies our agent design method is verified.

1. INTRODUCTION

Bidding strategies are the main factors in determining an agent's profit. A fare amount of research works can be found in the literature on this subject, that some are reviewed in the following. There has been three main categories in these researches: optimization-based approaches, equilibrium-based approaches and learning-based approaches.

In the first category, all the market environment and other agents are modeled stochastically or deterministically and the optimal bidding strategy of the agent is determined accordingly. For instance, in (Conejo et al, 2002) the probability density function of the next day hourly market clearing price (MCP) is estimated and a self-scheduling profit maximization process is obtained. In (Peng and Tomsovic, 2003) a Cournot model is applied to the bidding strategy problem. The bidding process together with the congestion management is modeled as a three level optimization problem. The congestion effect is explicitly formulated in the profit function. In (Ni and Luh, 2002) the bidding risk and self-scheduling requirements are managed by an optimization algorithm. In (Gross and Finaly, 1996), an analytical approach for building the optimal bidding strategy in the electricity market was developed under the assumption of a perfectly competitive market. In (Ma et al, 2005), a probability distribution for market clearing price (MCP) and parameters of other competitor's supply functions is obtained. Defining the profit as a function of MCP and other competitor's parameters, it is then optimized randomly with network constraints. The contribution of this article is not only the optimization of the profit function but also to minimize estimation error by entering variance term of the estimated profit function into the target function. The calculations are done using GA numerical methods. In the similar method introduced in (Rodriguez and Anders, 2004) predicted error distribution is identified in addition to price prediction, setting bidding steps according to how to treat the risk. There are mainly two approaches, risk aversion and risk taking. Risk taking agent prefers more profit with low probability to low profit with high probability and risk averse vice versa. In result, the maximum profit is made by risk aversion algorithm but risk taking agent unintentionally makes more profit when residual demand is low.

In the second category of researches, the rivals are not modeled into the environment but they are considered in determining game theoretic equilibrium of the market. In (Ferrero et al 1997) and (Park et al, 2001), game theory is applied to find the Nash equilibrium of the bidding game, corresponding to the optimal biddings of the participants.

Finally in the third category of bidding strategies, learning algorithms are applied to the bidding strategy problem. Due to the complexity of electricity markets, learning agents have been claimed to be more effective. In (Richter Jr et al 1999) and (Richter Jr and Sheble, 1998), evolutionary and artificial intelligence techniques such as genetic algorithms, genetic programming and finite state automata are used to develop adaptive and evolutionary bidding strategies. The drawback of the method is that all agents are adapting their strategies at each GA generation which makes it difficult to identify if a particular strategy is an appropriate one. In (Song et al, 2000), bidding strategy is represented as a multiple stage probabilistic decision-making problem and optimized by a Markov decision process. This model considers load and rival behavior uncertainty but ignores transmission constraints. In (Rahimi-Kian et al, 2005) a combination of fuzzy logic and learning is applied and the agent learns how to propose the

result of its fuzzy inference system considering the uncertainty in the quantity or price.

In (Rahimiyan and Mashhadi, 2007) two models are compared, one based on production optimization as a function of estimated MCP, and the other based on a Q-Learning which learns optimized production according to the past MCPs. It is concluded that the learning agent's behavior is more acceptable.

Although some methods estimate MCPs or rivals' behaviors, many researches focus on estimating residual demand (RD). RD is the effective demand observed by an agent, which can be calculated by subtracting opponents' bids from the market demand in economics. Knowing RD, it is shown in (Candiles et al, 2002) that an iterative algorithm leads to optimal bidding policy of a generating company. In (Mateo et al, 2000) genetic algorithm (GA) is applied to find the optimal strategy where each gene corresponds to one step of bidding. The fitness function is defined as the expected profit due to the other competitor's behavior, where a linearized probabilistic residual demand function is used to model other agents' uncertain behavior.

There have been some econometric approaches towards identification of residual demand function. In (Baker and Bresnahan, 1988), it is shown that residual demand function seen by an agent can be represented by following equation:

$P^*=RD(Q^*, Y, C, \tilde{C}; A, B, \Omega)$ where RD is the residual demand function, P^* and Q^* are price and quantity of the agent, Y is exogenous demand variables, C is the union of all specific factor vector of industry-wide factor price, $\tilde{C}$ is the vector of prices of all agents that are not in the industry-wide factor prices but may depend on some specification of the agents, A represents all demand parameters of all agents such as own-price demand elasticity as well as cross-price elasticity, B is all cost parameters and Ω is the union of all the behavior vector variable determining $\partial Qi/\partial Qj$. The parameters A, B, Ω should be estimated on economic data but because of the limitation of accessibility to these data, their joint impact on the slope of residual demand curve is estimated in their work.

Although (Baker and Bresnahan, 1988) has simplified the identification process of residual demand function, it doesn't suit electricity markets according to their complexity, time varying properties and more limited access to market's data.

In this research, a combination of optimization and learning methods based on an estimation of RD is employed to find the optimal bidding strategy. In section two, the design of agent's mind will be explained. In section three, simulation environment is described and finally some concluding remarks are expressed.

2. AGENT DESIGN

As our agent has separate tasks which make up its bidding behaviors, we designed the brain of our agent as a society of autonomous agents working cooperatively in order to enhance the bidding outcomes. The structure of the agent's mind is shown in the Fig. 1. The rewards and punishments received from the environment (the obtained profits) are distributed among them based on their duties, conditions and actions. In the figure below, the blue lines stand for regular data transfer, the red lines for reward/punishments assignment of the inner agents and the gray dash dot lines for the global policy information flow.

In the agent's mind, there are several agents learning together. Following, each sub-agent inside the agent's brain and its implementation algorithm will be explained.

2.1 Identification section:

In this section, the residual demand function is identified by a locally linear neuro-fuzzy (LLNF) model. According to residual demand function's shape in real electricity markets (e.g. Wolak, 2002), we considered the following equations as the model of equations (1).

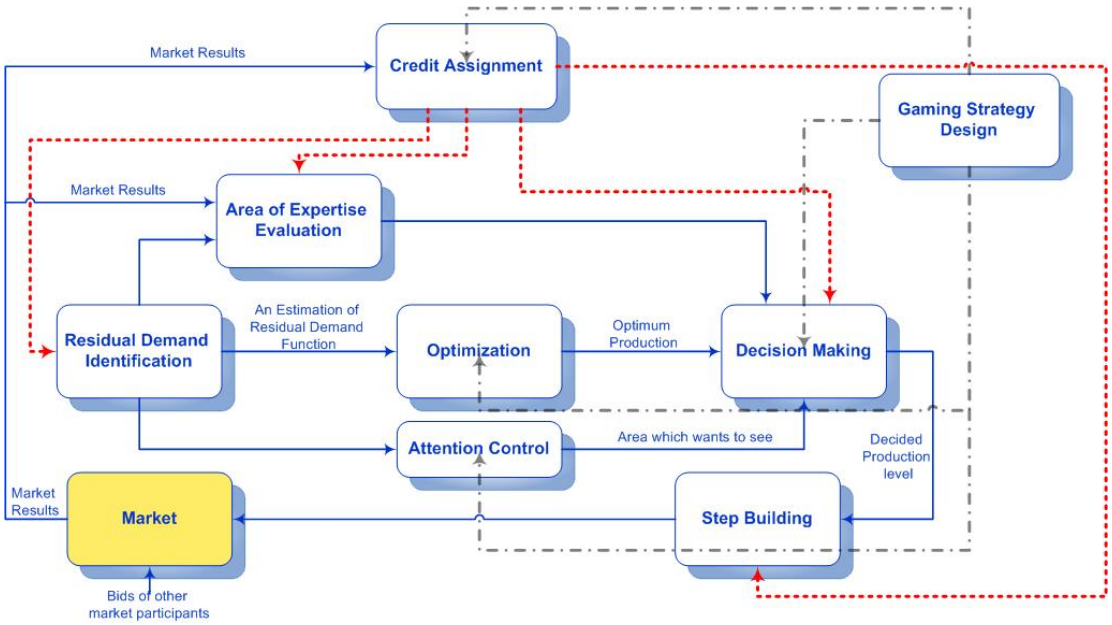

Fig.1 shows the structure of the designed agent's mind: In the agent mind there exists several agents learning together.

$$
\begin{aligned}
m_1 &= \exp\left[\left(\frac{q-M/8}{M/12}\right)^2\right], & \overline{m_1} &= \frac{m_1}{\sum m_i} \\
m_2 &= \exp\left[\left(\frac{q-M/2}{M/6}\right)^2\right], & \overline{m_2} &= \frac{m_2}{\sum m_i} \\
m_3 &= \exp\left[\left(\frac{q-7\times M/8}{M/12}\right)^2\right], & \overline{m_3} &= \frac{m_3}{\sum m_i} \\
p &= \sum_{i=1}^{3} \overline{m_i}\left(a_i q + b_i\right)
\end{aligned}
\tag{1}
$$

where price (p) is calculated according to production quantity (q). M is the maximum generation capacity. Parameters a_i and b_i are determined using Least Square methods based on the data history of the agent $\{p_i, q_i, t_i\}$, i.e. observing the price p_i while producing q_i at the time t_i. For better identification, greater weights are applied to the most recent data using a forget factor of γ=0.95.

2.2 Attention control section

Fatemi and Ahmadabadi (Fatemi and Ahmadabadi, 2007) proposed a general method for attention control in concept learning. In the first phase of their research, attentions are modeled as actions and attention control is learned with reinforcement learning. They conclude that their agent learns how to control its attention in order to recognize concepts. Although in our case, we are not looking for concepts in the market, we use their learning method of attention control.

With a close cooperation with the identification section, attention control section determines which parts of residual demand curve need attention. Our criterion for determining these parts is the scarcity of recent visits. Therefore Effective Number of Visits (ENV) is calculated by accumulating the forgetting factors around a specific point using a moving window of width 20. Using the history data $\{p_i, q_i, t_i\}$, ENV of a specific q at the time t is calculated via equation 2:

$$
ENV(q,t) = \sum_{q_i \in [q-10, q+10]} \gamma^{t-t_i} \tag{2}
$$

where γ is the forget factor (0.95). Finally at each time, t, the q which minimizes ENV(q,t) is selected as the block's output. Hence the question *"where to look in the environment"* is answered. Learning *"when to look"* is done in decision making section.

2.3 Expertise evaluation section

This section determines how expert the agent is in different parts of the residual demand function. Ahmadabadi and Asadpour (Ahmadabadi and Asadpour, 2002) proposed some indexes for determining of the Area Of Expertise (AOE). The basic use of AOE is in cooperative learning where an agent wants to select the most expert agent in specific states to learn from. They proposed "when rewards and punishments are approximately equal, use of *Abs* and *Nrm* indexes for determining AOE provide better results." We used Abs index, the sum of absolute values of rewards and punishments, in this research as it seems to be more efficient. Implementing this idea, a moving average window is applied on absolute values of estimation errors at each time. The errors are weighted according to the forget factor and as the little sum of absolute errors may be due to a lack of data, the moving average is normalized by number of points available in each region. To summarize, Expertness in the demand function around production level (q) at the time (t) can be calculated as equation 3:

$$
Expertness(q,t) = \frac{\sum_{q_i \in [q-10, q+10]} |e_i| \gamma^{t-t_i}}{\sum_{q_i \in [q-10, q+10]} 1} \tag{3}
$$

where e_i is the measured error at the time t_i when q_i , p_i were observed. It shall be noted that e_i is calculated based on the model made at the time t_{i-1} not at the time t.

According to the fact that expertness is a relative concept, the absolute values of Expertness(q,t) are not useful. Therefore they are converted a 3-valued Expertness Level (EL). At each time, we sort the values of expertness and assign the first (20+b)% EL of 1, the next (30+2b)% EL of 0 and the rest will be given EL of -1, where b is the bias learned by this section. (Learning the bias is explained in Credit Assignment section.)

2.3 Optimization section

In this section, a numerical optimization method searches for the optimum (profit maximizing) production quantity(q) according to the estimated residual demand function provided by equation 1.The profit maximization can be formulated as:

$$
\begin{aligned}
&\max_q \Pi(q) = pq - C(q) \\
&p = \sum_{i=1}^{3} \overline{m_i}\left(a_i q + b_i\right) \\
&0 < q < Maxq
\end{aligned}
\tag{4}
$$

where π is the profit, p is calculated using equation 1, C(q) is the cost function of the agent and Maxq is its maximum generation capacity.

2.4 Decision making section

This section determines the amount of planned production; this amount can be either the amount suggested by the optimizer section (Qo) or by the attention control section (Qa). The decision making depends on the expertise level (EL) in the residual demand function around Qo and Qa. The learning process of decision making is explained in the Credit Assignment section.

2.5 Step making

As the biding to the market should be proposed in an ascending step format, in this section making bidding steps for achieving the optimum production amount should be learned. Before each bidding, the goal production quantity (q*) is determined by decision making section and the estimated price (p*) according to q*, can be calculated from equation 1. Three step-making methods are considered. The best choice among these three shall be learned by this section.

In the first step-making method, we intend to intersect the residual demand by our bid steps vertically at the point (q*, p*). So quantities bellow q* are bided with the price of marginal cost and quantities more than that are bided with the market price cap. In the second method the residual demand function is aimed to be intersected horizontally. So quantities bellow 0.9q* are bided with the price of marginal cost, quantities more than 1.1q* are bided with the market price cap and the quantities between 0.9q* and 1.1q* are bided at the price of p*. Finally in the third method, we tend to intersect the residual demand curve diagonally. Therefore we bid quantities less than 0.9q* at the price of 0.9p* and quantities more than 1.1q* at the price of 1.1p*. The space between these two points is filled with equivalently distributed small steps.

2.6 Credit assignment

This section distributes the attained reward/punishment among the subagents. The identification section, independent of all made decision, is punished by its estimation error in price according to the resulted production of the market. In other words, identification section updates its estimates of a_is and b_is in equation 1 according to the new observations.

Similarly, the expertise evaluation section receives the absolute errors of identification section (e_i) and changes its output according to equation 3. It also receives a fixed amount of punishment/reward in the cases that it determines the agent as an expert one while it isn't expert or vice versa in order to amend its bias. The reward/punishment is determined based on table 1. To learn this bias (b), a Q-Learning has been implemented which has five states for different values of b: {0, 5, 10, 15, and 20} and three actions which are {increasing, decreasing or making no change in b}. The Q values are updated by traditional method:

$$Q(s,a) \leftarrow Q(s,a) \times (1-\alpha) + \alpha \times reward \tag{5}$$

where reward is determined by table 1 and α is considered 0.1 in our implementation.

The reward signal for the decision making part is the difference of the previous stage profit (π) and the average profit of the previous 10 stages (Aveπ). The correct choice between the inputs is learned via a Q-learning where the states are a combination of the history of the actions and the expertness level (EL) in the area of inputs:

Table1: Expertness Evaluation section's reward determination

Reward (+) / Punishment (-)		Expertness Level (EL)		
		1	0	-1
Relative Error	<10	1	0	-1
	>10 & <30	0	1	0
	>30	-1	0	1

$$Q(s,a) \leftarrow Q(s,a) * (1-\alpha) + \alpha(\Pi - Ave\Pi)$$
$$s = \{a_{t-1}, a_{t-2}, EL(input1), EL(input2)\} \tag{6}$$

α is considered 0.1 in our implementation and input1 and input2 are outputs of optimization block and attention control block respectively.

The step-making part would be punished according to absolute difference between production amount ($q_{produced}$) and decision making ordered amount ($q_{ordered}$). This section employs Q-Learning too, but has only one state. The actions are the three choices for step-making. Hence the update method would be:

$$Q(a) \leftarrow Q(a) * (1-\alpha) + \alpha \times \left| q_{produced} - q_{ordered} \right|, a \in \{1,2,3\} \tag{7}$$

α is considered 0.1 again.

2.7 Gaming strategy section

In this section the main policy of the bidding would be determined. According to current researches in multi agent reinforcement learning (Shoham et al, 2003), acting passively in a game does not always assure good result. In electricity market words, an agent must not always believe in itself as a price taker. Therefore after reaching a relative stability the agent should change its strategy according to its influence on the market (its effect on its residual demand curve). The effects of these strategies are shown in figure1 using gray dash dot lines.

Although this part is left as the future works of this research, it should be noted that for the first 100 rounds of the market, the strategy is to explore its environment in the simulation. Some further explanations are provided in the second simulation.

3. SIMULATION AND RESULTS

In order to verify our approach, an agent with the mentioned descriptions is made and its performance is simulated in a multi agent environment. There are some other agents as the rivals in this environment and their descriptions are summarized in table 2. For simplicity no network constraint were considered in our simulation. For all players, the cost function is $C(q)=0.05q^2+10q+100$, the generation capacity is 400 and market's price cap is 100. The market is a uniform price one.

In the first experiment, our designed agent competes against SWR agent. The market demand in each round is between 450 and 500 randomly. The results illustrated in fig. 2 are surprising. As it can be observed, our agent's profit is less than SWR's; but the point is that SWR is bidding for a fixed amount of production which is not necessarily the optimum one. Our agent's strategy is to bid for little production so that market price reaches its cap and our agents profit is maximized. However when the market price reaches its cap, SWR is capturing the majority of the market production and therefore gains more profit than ours. Comparison between our agent and other mentioned agents' profits against SWR would indicate the performance of our strategy.

Table2: Different agents of the multi agent environment

Agent Name	Category	Description
SWR (Satisfied Without Risk)	Non-Learning	This agent would bid in order to cover its costs for sure. It divides its capacity into n (in our case n=4) sections and in each section it bids 10% over its marginal cost.
SLR (Satisfied Low Risk)	Non-Learning	This agent is similar to the previous agent but it randomly bids something more than 10% over its marginal cost.
OPL (Overcharging Percentage Learning)	Learning	This agent is similar to the SLR agents, but it also tries to learn overcharging percentage with a simple Q-learning.
PL (Price Learning)	Learning	This agent adaptively estimates average market price (P*) and finds its optimal production (Q*) in this price using a simple profit maximization. Then it proposes market price cap for the amounts higher than Q* they bid and its marginal cost for the lower quantities. In this way it would produce its managed power (Q*) in the market.

Simulations show that our agents, PL, OPL and SLR in competition with SWR obtain about 3700, 1700, 2000 and 1500 units of profit respectively. It's also deduced that in a uniform price market, bidding low prices in order to capture the production is a good strategy as other agents would probably bid in a way that market clearing price becomes high.

In the second experiment our agent is competing OPL and demand level is lower than each producer's capacity. In this condition the residual demand curve would shift towards the vertical axis and the price would never reach to its market price cap. In this experiment our agent changed its bidding algorithm according to the information of the first experiment and after 100 rounds of bidding; from rounds 101 to 500 lowered its bids to encourage its competitor to bid higher. During these rounds our agent's production increased because of lower proposed prices, but since its competitor is not intelligent enough to increase the market price considerably; our agent's profit is not acceptable although it is higher than its opponent.

During the rounds 500 to 1000 our agent bids in a way that the market clearing price is increased. Although the competitor benefits more from this price shift, our agent's profit is also increased. Again it is seen that having more profit than the rival is not the goal, the goal is to maximize our profit regardless of others. In this experiment a task of gaming strategy design section exists in the rounds 100 to 500.

In the last experiment, all agents are put in a random demand environment. Although having the most profit is not the criteria of our agent's performance, it can be seen that our agent profit is the highest among other competitors. (Fig. 4)

4. CONCLUSION

In this paper a mind structure for a bidding agent in electricity markets was designed. The agent's brain consisted of cooperative sub-agents that learned together for better competency of the main agent. Each subagent had a specific task and a credit assignment unit was designed in order to distribute the rewards/punishment among the learning subagents. The agent's expected profit maximization together with identification of the key market factors (such as the residual demand) and its expertise (by learning in the market) could help it perform better than others in the market. This combination of different criteria for performance evaluation of an agent in the market environment was the main contribution of this paper compared to the single criterion evaluation in other research works.

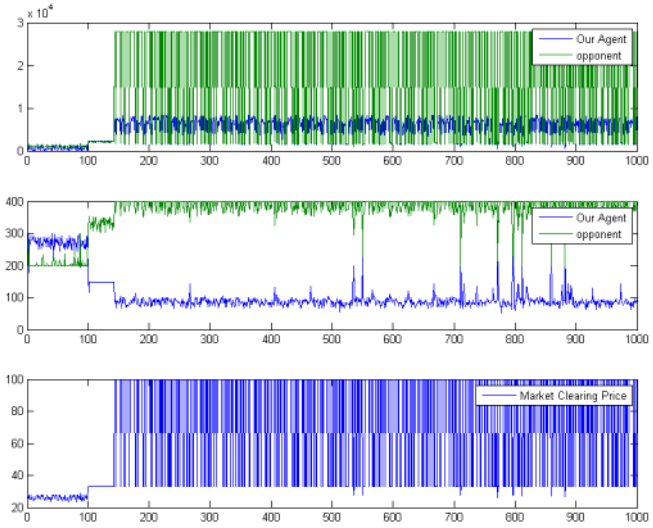

Fig2: Competition between the designed agent and SWL, Top to bottom: Profit, production, market price diagram

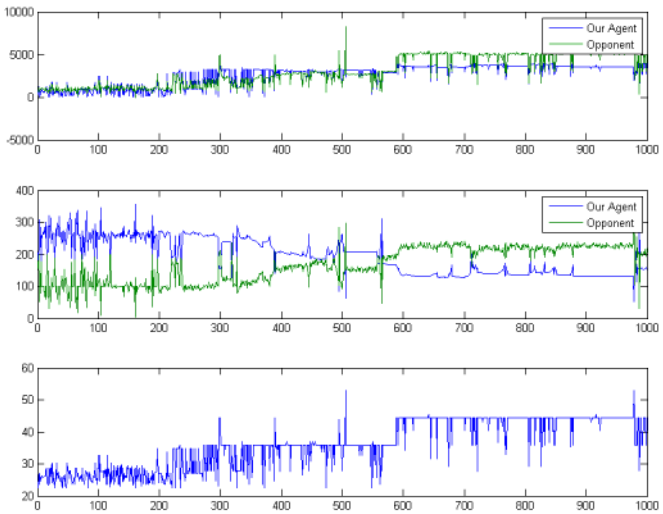

Fig. 3: Competition between our agent and OPL. Top to bottom: Profit, production, market price diagram

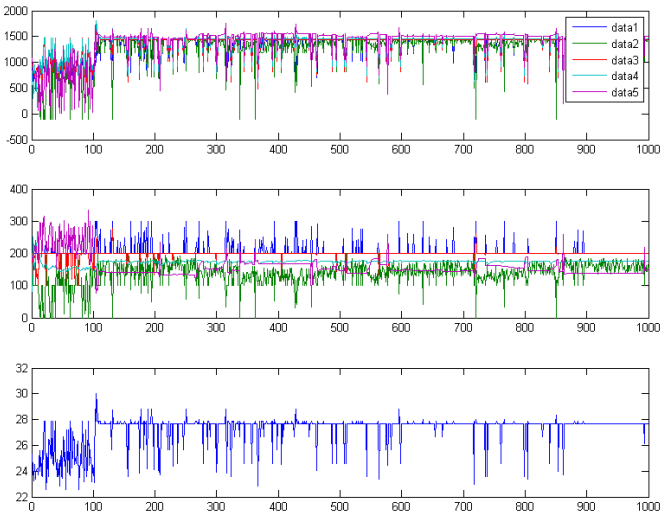

Fig.4: Competition among all agents. Top to bottom: Profit, production, market price diagram Data 1: SWL (blue), data2: OPL (green), data3: SLR (red), data 4: PL (Cyan), data5: our designed agent (Magenta)

By analyzing the simulation results, firstly we observed that in the competition of two agents, the one who attains more profit than the other does not necessarily have better strategies. Our best strategy against a specific agent is the one that maximizes our profit regardless of the fact that our profit is more or less than the competitor's profit. Secondly, we deduced that the successfulness of changing the gaming strategies depends strongly on the opponent's intelligence. Competing with an intelligent rival is usually more beneficial in electricity markets.

A multi-agent environment was simulated for evaluating the designed agent's performance. The simulation results showed a promising performance of our agent compared to other agents with different designs.

REFERENCES

Ahmadabadi, M.N., M. Asadpour (2002). Expertness Based Cooperative Q-Learning. *IEEE Trans. on Systems, Man, And Cybernetics—Part B: Cybernetics*, **32(1)**, 66-76.

Baker, J. B. and T. F. Bresnahan (1988). Estimating the Residual Demand Curve Facing A Single Firm. *International Journal of Industrial Organization*, **6**, 283-300.

Candiles, J. O., J. I. de la Fuente and T. Gómez (2002). A cobweb bidding model for competitive electricity markets. *IEEE Trans. Power Syst.*, **17**, pp. 148–153.

Conejo, A. J., F. J. Nogales and J. M. Arroyo (2002). Price-taker bidding strategy under price uncertainty. *IEEE Trans. Power Syst.*, **17**,1081–1088.

Fatemi, H., M.N. Ahmadabadi, (2007). Biologically Inspired Framework for Learning and Abstract Representation of Attention Control, In press.

Ferrero, R. W., S. M. Shahidehpour and V. C. Ramesh, (1997). Transaction analysis in deregulated power systems using game theory. *IEEE Trans. Power Syst.*,**12**, 1340-1347.

Gross, G. and D. J. Finlay (1996). Optimal bidding strategies in competitive electricity markets. *In Proc. 12th Power Syst. Computat. Conf. (PSCC'96)*, 815–823.

Ma, X, F Wen, Y Ni and J Liu (2005). Towards the development of risk-constrained optimal bidding strategies for generation companies in electricity markets. *Electric power systems research*, **73(3)**, 305-312.

Mateo A., E. F. Sánchez-Úbeda, A. Muñoz, J. Villar, A. Saiz, J.T. Abarca, E. Losada (2000). Strategic Bidding under Uncertainty using Genetic Algorithm, *PMAPS Conference ,Madeira, Portugal, September, 2000*.

Ni, E. and P. B. Luh (2002). Optimal integrated generation bidding and scheduling with risk management under a deregulated daily power market. *In Proc. 2002 IEEE-Power Eng. Soc. Winter Meeting*, New York.

Park, J.-B., B. H. Kim, M.-H. Jung and J.-K. Park (2001). A continuous strategy game for power transactions analysis in competitive electricity markets. *IEEE Trans. Power Syst.*,**16**, 847–855.

Peng, T. and K. Tomsovic (2003). Congestion Influence on Bidding Strategies in an Electricity Market. *IEEE Trans. Power Syst.*, **18(3)**, 1054-1061.

Rahimi-Kian, A., B. Sadeghi and R.J. Thomas (2005). Reinforcement Learning Based Supplier-Agents for Electricity Markets. *Proceedings of the Power Engineering Society General Meeting 2005*, pp. 420-427.

Rahimiyan, M. and H. R. Mashhadi, Supplier's optimal bidding strategy in electricity pay-as-bid auction: Comparison of the Q-learning and a model-based approach. *Electric Power Systems Research*, In Press, Available online 26 March 2007.

Richter Jr., C. W. and G. B. Sheblé, (1998). Genetic algorithm evolution of utility bidding strategies for the competitive marketplace, *IEEE Trans. Power Syst.*, **13**, 256–261.

Richter Jr, C.W., G. B. Sheblé and D. Ashlock (1999). Comprehensive bidding strategies with Genetic programming/finite state automata. *IEEE Trans. Power Syst.*, **14(4)**, 1207–1212.

Rodriguez, C.P. and G.J. Anders (2004). Bidding Strategy Design for Different Types of Electric Power Market Participants. *IEEE Trans. Power Syst.*, **19(2)**, 964- 971.

Shoham, Y., R. Powers and T. Grenager, (2003). Multi Agent Reinforcement Learning: a critical survey, Technical report, *Stanford University*, url: citeseer.ist.psu.edu/shoham03multiagent.html.

Song, H., C.-C. Liu, J. Lawarrée, and R. W. Dahlgren (2000). Optimal electricity supply bidding by Markov decision process, *IEEE Trans. Power Syst.*, **15**, 618–624.

Wolak F., (2002). *Identification and Estimation of Cost Functions Using Observed Bid Data: An Application to Electricity Markets*, Web Site: http://www.stanford.edu/~wolak, 2002.