Agent-Based Modeling of Day-Ahead Real Time Pricing in a Pool-Based Electricity Market

Sh. Yousefi*, M. Parsa Moghaddam* and V. Johari Majd*

Abstract: In this paper, an agent-based structure of the electricity retail market is presented based on which day-ahead (DA) energy procurement for customers is modeled. Here, we focus on operation of only one Retail Energy Provider (REP) agent who purchases energy from DA pool-based wholesale market and offers DA real time tariffs to a group of its customers. As a model of customer response to the offered real time prices, an hourly acceptance function is proposed in order to represent the hourly changes in the customer's effective demand according to the prices. Here, Q-learning (QL) approach is applied in day-ahead real time pricing for the customers enabling the REP agent to discover which price yields the most benefit through a trial-and-error search. Numerical studies are presented based on New England day-ahead market data which include comparing the results of RTP based on QL approach with that of genetic-based pricing.

Keywords: Day-ahead Real-time Pricing, Genetic Algorithm, Hourly Acceptance Function, Multi-Agent Systems, Q-Learning.

1 Introduction

Sophisticated processes in mutual relations among different decision makers in a retail market provide enough justification to model the market using the concept of agent theory. A wide range of agent theory applications have been reported in power engineering studies from long term planning to real time operation [1-6]. Ref. [1] has simulated customer agents' decision making in switching their corresponding supplier once a year. A set of trading strategies in a multi-agent energy market composed of autonomous computational agents assigned to the suppliers, retailers and customers has been addressed in [2]. Ref. [6] has presented a general model of the interaction among competitor retailers and heterogeneous consumers using a multi agent system (MAS), in which the retailers take time-of-use (TOU) pricing strategy. In this paper, an agent-based structure of the electricity retail market is presented based on which day-ahead (DA) energy procurement for different clusters of customers is modeled. The presented multi-agent competitive retail electricity market is applicable in almost all retail market studies and allows admitting new members such as new REPs and customers due to the inherent modularity characteristic of MAS.

Electricity markets have been focused by many researchers in recent years [7-10]. Customers' participation in demand response programs [11-14] and pricing strategies in power markets [14-16] are some related subjects of research.

In an intelligent power grid, customers have the right to switch to their desirable energy suppliers corresponding to the different offered pricing schemes and demand response programs even in a short-term market such as DA one. Ref. [9] has focused on fixed pricing pattern modeling the customer's behavior in selecting the most profitable retailer among some competitor REPs. It has utilized a type of market share function which shows the percentage of the overall load that may be demanded by the customers at different fixed prices. A linear acceptance function has been proposed in ref. [15] specifying the probability of accepting the offered TOU prices by the end-users. Ref. [16] has proposed a cluster-based acceptance function which shows the maximum acceptable energy costs for different customers. Based on this model, if the offered prices result in energy costs beyond acceptable thresholds, the customers reject and purchase their whole electricity demand from other retailers. The latter model does not represent the percentage of the load which customers may decrease or purchase from other retailers according to the offered prices.

Iranian Journal of Electrical & Electronic Engineering, 2011.
Paper first received 22 Jan. 2011 and in revised form 3 July 2011.
* The Authors are with the Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran.
E-mails: sh.yousefi@modares.ac.ir, parsa@modares.ac.ir and majd@modares.ac.ir

This paper models energy procurement by one of the REP agents for three clusters of its customers according to fixed, TOU and real time pricing patterns. Hereinafter, the customer agents representing customers who adopt fixed, TOU and real time (RT) tariffs are referred to as fixed, TOU and RT price agents, respectively. The REP agent gains from the pool-based wholesale market and procures electricity for its customers. Real time pricing for RT price agent is addressed here based on learning capabilities of intelligent agents. The REP agent's intelligence appears in the ability of learning the optimal pricing strategy by experiencing its impact on RT agent's demands and consequent retail profit. REP agent's learning capability is modeled based on Q-learning (QL) approach. The REP agent must discover which price yields the most benefit through a trial-and-error search. Specifically, a daily 24-h horizon for energy procurement and real time pricing (RTP) is considered. The REP agent is able to forecast energy consumption of fixed and TOU price agents using heuristic methods like Artificial Neural Networks (ANNs) so as to decide on the amount of electricity to purchase in DA market. As in a competitive retail market, the procedure addressed here allows the customers belonging to RT agent to adjust their effective demand according to the real time retail prices and/or change their retail energy provider even in a daily basis. An hourly acceptance function (HAF) composed of 24 adjusted functions is proposed in this paper in order to represent demand-side response to the offered RT rates.

The remaining parts of the paper are organized as follows. Section 2 presents the agent-based retail market model and its intelligent agents. Section 3 is on the energy procurement for customers. Section 4 represents the proposed model of customers' demand function. Section 5 is devoted to real time pricing for the active customers. Section 6 is assigned to the numerical studies. Finally, section 7 concludes the paper.

The main notation used throughout the paper is stated below for quick reference. Other symbols are defined as required throughout the text.

$\mathrm{AC}(\mathrm{P_{RT}}(\mathrm{h}))$	Acceptance of the real time price $\mathrm{P_{RT}}$
$\mathrm{B}(\mathrm{P_{RT}}(\mathrm{h}))$	Benefit resulted from selling electricity to the customers at $\mathrm{h}-\mathrm{th}$ hour
$\mathrm{D^{in}}(\mathrm{h})$	Surplus energy need for inactive customer agents (MWh)
$\mathrm{D^{a}}(\mathrm{P_{RT}}(\mathrm{h}))$	DA demand of active customers at $\mathrm{h}-\mathrm{th}$ hour (MWh)
$\mathrm{P_{RT}}(\mathrm{h})$	Offered real time tariff at $\mathrm{h}-\mathrm{th}$ hour ($/MWh)
$\mathrm{Q}(\mathrm{s_i},\mathrm{a_i})$	Action-value function of action a_i in state s_i

A superscript i affecting any of the symbols above indicates the related value at the $\mathrm{i}-\mathrm{th}$ iteration of Q-learning process.

2 Retail Market as a Multi-Agent System

Considering a pool-based wholesale market, the structure of an agent-based retail environment is depicted in Fig. 1 including defined types of heterogeneous communicating agents.

In the structured multi-agent retail market model (Fig. 1), there are as competitive REP agents as retail energy providers in a retail environment. The REP agent, as the main intelligent agent in this study has the ability of learning the optimal strategy which is modeled using Q-learning approach.

From the REP agent's viewpoint, customers are categorized corresponding to their adopted pricing patterns. As it can been seen in Fig. 1, in this stage of our research, the customers are classified in three groups of fixed, TOU and RT price agents. Fixed and TOU price agents, also called here as inactive customer agents, wish access to reliable, qualified, and inexpensive electricity but not engaging in price variations. These customer agents do not need to determine the exact amount of their hourly demand because of the inherent nature of the adopted contracts. RT price agent, also called here as active customer agent, represents active customers who usually purchase a portion of their demand through bilateral contracts and then bid in day-ahead and spot markets to procure their surplus electric power need aiming at managing their electricity bills. Active customers monitor other pricing alternatives and different provided services and have a good prediction about their hourly demand. Distributed Energy Resources (DERs) refer to Distributed Generators (DGs) and energy storage systems such as batteries and electric vehicles. In this paper, a pool-based wholesale market is considered in which different types of large scale power plants and DERs sell their energy output at the market cleared price. According to the point of energy delivery, the REP agent purchases electricity from the wholesale market at hourly Locational Marginal Prices (LMP). As it is shown in the shaded part of Fig. 1, the focus of this paper is on the retailing interrelations between customer agents and their relevant REP agent which procures energy from the pool-based wholesale market at known hourly LMP ($P_w(h)$). The REP agent has an indirect competition interaction with other REPs via customer response to the offered price appeared in an hourly acceptance function.

3 Energy procurement for customers

A retail energy provider procures most of its electricity needs via long-term contracts based on its knowledge. During days approaching the exact time of power delivery, the REP agent gradually estimates the enrolled customers' demand and updates its knowledge base, then procures the remaining portion of required power in DA and real time markets from the wholesale market. The REP agent applies forecasting approaches to predict probable variations in real time consumption

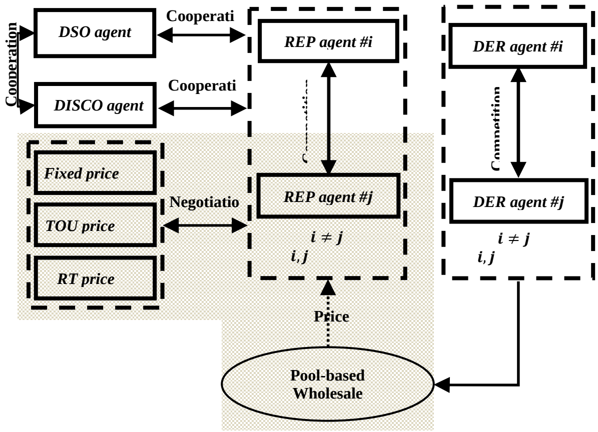

Fig. 1 Agent-based structure of the electricity retail market.

of the fixed and TOU price agents and updates its knowledge base, subsequently. ANNs are one of the most widely used approaches among load forecasting methods which attempt to discover mathematical links between inputs and outputs and are able to adapt themselves to rapid changes in load profile. Here, 24 multilayer neural networks (for 24 hours of the next day) are designed and Bayesian regularization approach is used as the networks' training function. Each neural network is composed of three layers, namely input layer with 2 neurons (DA demand and the next day's forecasted temperature), a hidden layer with 40 neurons and one neuron output layer.

It is considered that a portion of previously predicted consumption of the inactive customer agents, typically 0.9 of it, has been supplied in long-term contracts. The surplus energy need for inactive customer agents is denoted by $\mathrm{D^{in}(h)}$ which can be obtained from the prior forecast of inactive agents' hourly demand ($\mathrm{D_1^{in}(h)}$) and recent forecast of it ($\mathrm{D_2^{in}(h)}$) as it is shown by Eq. (1).

$$\mathrm{D^{in}(h)} = 0.1 \times \mathrm{D_1^{in}(h)} + \frac{1+\mathrm{sign}\left(\mathrm{D_2^{in}(h)} - \mathrm{D_1^{in}(h)}\right)}{2} \times \left(\mathrm{D_2^{in}(h)} - \mathrm{D_1^{in}(h)}\right) \tag{1a}$$

where,

$$\mathrm{D^{in}(h)} = \mathrm{Fix(h)} + \mathrm{TOU(h)} \tag{1b}$$

$$\mathrm{D_1^{in}(h)} = \mathrm{Fix_1(h)} + \mathrm{TOU_1(h)} \tag{1c}$$

$$\mathrm{D_2^{in}(h)} = \mathrm{Fix_2(h)} + \mathrm{TOU_2(h)} \tag{1d}$$

The surplus energy need for fixed and TOU price agents are represented by $\mathrm{Fix(h)}$ and $\mathrm{TOU(h)}$, respectively and h denotes the target hour of the next day. Also, $\mathrm{Fix_1(h)}$ and $\mathrm{TOU_1(h)}$ represent prior forecasts of fixed and TOU price agents' demands at $\mathrm{h-th}$ hour of the next day, respectively. Similarly, $\mathrm{Fix_2(h)}$ and $\mathrm{TOU_2(h)}$ represent recent forecasts of fixed and TOU price agents' demands at the same hour, respectively.

In order to supply the active customers with electricity of hourly varying price, the REP agent proposes DA prices and experiences subsequent customers' reactions by observing active agent's demand as a function of the offered RT tariffs denoted by $\mathrm{D^a}\left(\mathrm{P_{RT}(h)}\right)$. Therefore, the amount of energy to be purchased in DA market by the REP agent ($\mathrm{D(h)}$) is as the following:

$$\mathrm{D(h)} = \mathrm{D^{in}(h)} + \mathrm{D^a}\left(\mathrm{P_{RT}(h)}\right) \tag{2}$$

An hourly acceptance function is proposed here to model active customers' response to the offered RT prices. This model is based on the market share function addressed in [9] which represents the acceptance of fixed prices by the customers.

4 Modeling Customers' Demand Function

The probability density function according to which the retailer sets its equilibrium prices is usually defined as the market share or acceptance function [9]. The acceptance function is a decreasing function in which by increasing the offered price, the total confirmed demand will be decreased. Several factors such as long-term strategy of the retailer, risk attitude, the behavior of customers and the behavior of the competitors must be considered in determination of the acceptance function [9,17]. Ref. [9] has applied a market share function (MSF) similar to Eq. (3) as the acceptance function for fixed pricing$\left(\mathrm{AC(P_{fix})}\right)$ which shows the percentage of the overall load that may be demanded by the retailer's customers at different fixed prices (P_{fix}).

$$\mathrm{AC(P_{fix})} = 1 - \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mathrm{P_{fix}}} e^{-(0.5)(\frac{t-m}{\sigma})^2} dt \tag{3}$$

The cost of energy procurement changes hourly and influences the offered prices by the REP and the customers' reaction against the offered prices. Due to the fact that DA real time pricing consists of 24 fixed-price time periods, an hourly acceptance function (HAF) is structured here which includes a set of 24 adjusted functions. To adjust hourly functions, it may be useful to focus on the descending trend of fixed-prices acceptance function. As illustrated by Eq. (4), the price at decreasing point of the above acceptance function (dp) is the highest offered price in which $AC\left(P_{fix}\right)$ equals one with a given tolerance (Fig. 2).

$$\mathrm{dp} = \{\mathrm{P_{fix}} | \mathrm{AC(P_{fix})} = 1 \,\&\, \mathrm{AC(P_{fix}} + \varepsilon) < 1\} \tag{4}$$

The functions are adjusted by setting different decreasing points aiming at proper modeling of customers' responses to the hourly prices ($\mathrm{P_{RT}(h)}$). Considering a marginal benefit for the retailers, the prices at hourly decreasing points ($\mathrm{DP(h)}$) are proportional to hourly DA energy prices offered by the wholesale market ($\mathrm{P_w(h)}$), as shown in Eq. (5).

$$\mathrm{DP(h)} = \mathrm{c} + \mathrm{P_w(h)} \tag{5}$$

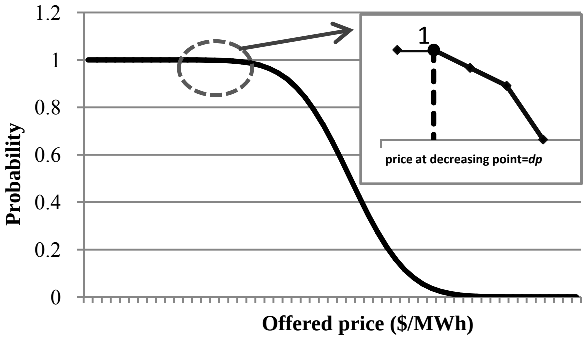

Fig. 2 Acceptance function for fixed pricing.

The constant c is influenced by some factors such as the local price caps and other REPs' pricing strategies. Setting the decreasing points as Eq. (5), results in shifting hourly price profiles accordingly. Equation (6) illustrates the proposed HAF.

$$\mathrm{AC}\big(\mathrm{P_{RT}(h)}\big) = 1 - \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mathrm{P_{RT}(h)}} e^{-(0.5)(\frac{t+dp-DP(h)-m}{\sigma})^2} dt \quad (6)$$

As it is shown in Fig. 3, HAF demonstrates the active customers' rational behaviors in forms of price acceptance values in the range of [0,1] for each hour of RTP. Fig. 3 is extracted based on a typical curve of DA wholesale prices which is applied in determining decreasing points of the acceptance functions.

The proposed HAF leads to a new type of demand vs. price function as represented by Eq. (7).

$$\mathrm{D^a}\big(\mathrm{P_{RT}(h)}\big) = \mathrm{D_0^a(h)AC}\big(\mathrm{P_{RT}(h)}\big) = \mathrm{D_0^a(h)} - \frac{\mathrm{D_0^a(h)}}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mathrm{P_{RT}(h)}} e^{-(0.5)(\frac{t+dp-DP(h)-m}{\sigma})^2} dt \quad (7)$$

Substituting Eq. (5) in Eq. (7) yields the proposed model of customers' demand as follows.

$$\mathrm{D^a}\big(\mathrm{P_{RT}(h)}\big) = \mathrm{D_0^a(h)} - \frac{\mathrm{D_0^a(h)}}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mathrm{P_{RT}(h)}} e^{-(0.5)(\frac{t+dp-c-m-P_w(h)}{\sigma})^2} dt \quad (8)$$

The above model represents the relationship between the offered retail prices and the customers' effective demand which is influenced by the hourly LMP.

5 Real Time Pricing for the Active Customers

Based on Eq. (8) as the model of active agent's demand in response to the offered RT prices in DA retail market, the REP agent's obtained benefit is formulated as Eq. (9).

It is considered that the amount of energy which is trading in the wholesale market is much larger than the REP agent's requisite energy. This model is defined for each of the 24 hours of the next day. Note that in each hour, initial demands ($\mathrm{D_0^a(h)}$) are known and DA wholesale price is forecasted while RTP rate is the variable to be determined. Constraint (9c) establishes a floor for RTP rates while the impacts of hourly acceptance function on the adopted rates would limit the upper bound of prices.

$$\begin{aligned}\underset{\mathrm{P_{RT}(h)}}{\mathrm{Maximize}}\ \mathrm{B}\big(\mathrm{P_{RT}(h)}\big) &= \mathrm{Fix(h)[P_{Fix} - P_w(h)]} \\ &+ \mathrm{TOU(h)[P_{TOU}(h) - P_w(h)]} \\ &+ \left(\mathrm{D_0^a(h)} - \frac{\mathrm{D_0^a(h)}}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\mathrm{P_{RT}(h)}} e^{-(0.5)(\frac{t+dp-c-m-P_w(h)}{\sigma})^2} dt\right) \\ &\times \mathrm{[P_{RT}(h) - P_w(h)]}\end{aligned} \quad (9a)$$

where,

$$\mathrm{P_{TOU}(h)} = \begin{cases} \mathrm{p_{TOU(peak)}(h)} & \mathrm{h \in \{peak\}} \\ \mathrm{p_{TOU(off-peak)}(h)} & \mathrm{h \in \{off-peak\}} \end{cases} \quad (9b)$$

subject to:

$$\mathrm{P_{RT}(h) \geq P_w(h)} \quad (9c)$$

5.1 Q-Learning Approach in RTP and Energy Procurement for Customers

The REP does not know which price to select in order to maximize its benefit. Therefore, it acts as a learner agent and tries to discover which options yield the most subsequent rewards or penalties, gradually. The REP agent can learn from its past experienced strategies which can be computationally implemented by using a Q-Learning algorithm. In this study, one-stepQ-learning is applied as REP agent's learning approachso as to reach optimized benefit while satisfying active customers. Let $\mathrm{S} = \{\mathrm{s_1, s_2, \dots, s_{n_s}}\}$ be the finite set of possible states in RTP process and $\mathrm{A} = \{\mathrm{a_1, a_2, \dots, a_{n_a}}\}$ be the finite set of admissible actions the agent can take. At each time step $\mathrm{t_i}$, the agent senses the current state $s_i \in \mathrm{S}$ and on that basis selects an action $\mathrm{a_i = a} \in \mathrm{A}$. In each state, three possible actions are defined as represented by Eq. (10).

$$\mathrm{A} = \{\mathrm{a_+, a_-, a_0}\} \quad (10a)$$

where,

$$\mathrm{P_{RT_i}(h)} \xrightarrow{\mathrm{a_+}} \left(\mathrm{P_{RT_{i+1}}(h) = P_{RT_i}(h) + \Delta P}\right) \quad (10b)$$

$$\mathrm{P_{RT_i}(h)} \xrightarrow{\mathrm{a_-}} \left(\mathrm{P_{RT_{i+1}}(h) = P_{RT_i}(h) - \Delta P}\right) \quad (10c)$$

$$\mathrm{P_{RT_i}(h)} \xrightarrow{\mathrm{a_0}} \left(\mathrm{P_{RT_{i+1}}(h) = P_{RT_i}(h)}\right) \quad (10d)$$

where $\mathrm{P_{RT_i}}$ denotes the offered sale price at $\mathrm{h-th}$ hour to the active customers in the $\mathrm{i-th}$ iteration of learning procedure. Here, a combination of soft-max and greedy policies is applied in pricing procedure. In each stage, based on the adopted policy, an action is selected. The above mentioned policies are based on Boltzmann distribution (Eq. (11)) and maximum probabilities, respectively.

$$\mathrm{p(s_i, a_i, i)} = \frac{e^{\mathrm{Q_{i-1}(s_i,a_i)/T_i}}}{\sum_{j=1}^{\mathrm{n_a}} e^{\mathrm{Q_{i-1}(s_i,a_j)/T_i}}} \quad (11)$$

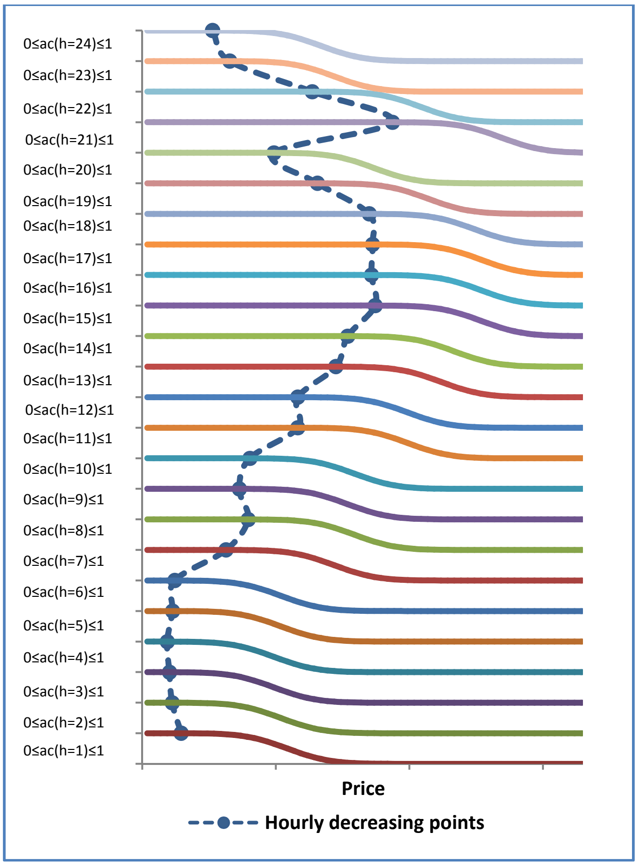

Fig. 3 HAF composed of adjusted acceptance functions based on hourly decreasing points curve.

where $i = 1,2,\cdots,L$ denotes the number of learning iteration and $p(s_i, a_i, i)$ represents the probability of selecting the action a_i related to the state s_i. The temperature T_i usually decreases during the learning iterations. Here, the reduction pattern is as the following equation:

$$T_i = L(1 - (i-1)/T_1) + 1e - 5 \tag{12}$$

where L denotes limited number of stages as the learning procedure's termination criterion. The updated price as a result of the adopted action affects the retailing benefit value and accordingly leads the agent to a new state of pricing strategy learning. In this study, three states are conceivable for the agent; $S = \{s_+, s_-, s_0\}$ including the states of gaining more benefit (s_+), less benefit (s_-) or no change (s_0) as represented by Eq. (13).

$$s_i = \begin{cases} s_+ = +1 & \text{if } B_i(P_{RT}(h)) > B_{i-1}(P_{RT}(h)) \\ s_0 = 0 & \text{if } B_i(P_{RT}(h)) = B_{i-1}(P_{RT}(h)) \\ s_- = -1 & \text{if } B_i(P_{RT}(h)) < B_{i-1}(P_{RT}(h)) \end{cases} \tag{13}$$

REP agent receives an immediate reward (r_i) which is proportional to the resulted change in its benefit value. Accordingly, the current state of the agent updates to the new state (s_{i+1}). Equation (14) represents the reward assigned to the action a_i from the old state s_i which has caused changes in REP agent's obtained benefit.

$$r_i = 100 \times s_{i+1} + 1e - 3 \times (s_{i+1} + 1) \tag{14}$$

The offered reward impresses action-state value function $Q(s_i, a_i)$ as represented by Eq. (15).

$$Q(s_i, a_i) = Q(s_i, a_i) + \alpha[r_i + \gamma \max_a Q(s_{i+1}, a) - Q(s_i, a_i)] \tag{15}$$

where α denotes the step-size parameter $(0 < \alpha < 1)$ and γ represents the discount rate in the range of [0,1]. These functions determine the most probable actions for the next play and are applied in taking the next action based on the mentioned policies. The offered price and the related benefit reach to their final values as the learning process terminates. The flowchart of the proposed Q-learning based method of real time pricing for the active customers is presented in Fig. 4.

5.2 Genetic Algorithm in RTP and Energy Procurement for Customers

In order to verify the results of QL-based RTP proposed in this paper, a comparison is made using genetic algorithm method. GA approach searches for an optimal price operating on populations of individuals. Each individual chromosome represents a particular selection of the real time tariff corresponding to each hour of the next day. The GA-based real time pricing is structured as follows:

a. The initial population of prices is randomly generated.
b. Based on the objective function (Eq. (9)) and its constraints, a measure of fitness is assigned to each of individuals to estimate the probability of selecting each individual (RT price value) in forming the next population.
c. Applying the genetic operators of crossover and mutation, a new population is produced employing the selected individuals.
d. Repeat the above stages to reach an ignorable difference in the objective function value.

Table 1 demonstrates the adopted choices the basic GA parameters in real time pricing for the RT price agent.

Table 1 The adopted choices in GA-based pricing.

GA parameters	Adopted options
Population size	20
Generations	500
Stall generation limit	100
Fitness scaling function	Proportional
Crossover function	Heuristic
Crossover fraction	0.8
Selection function	Roulette
Mutation function	Adaptive feasible

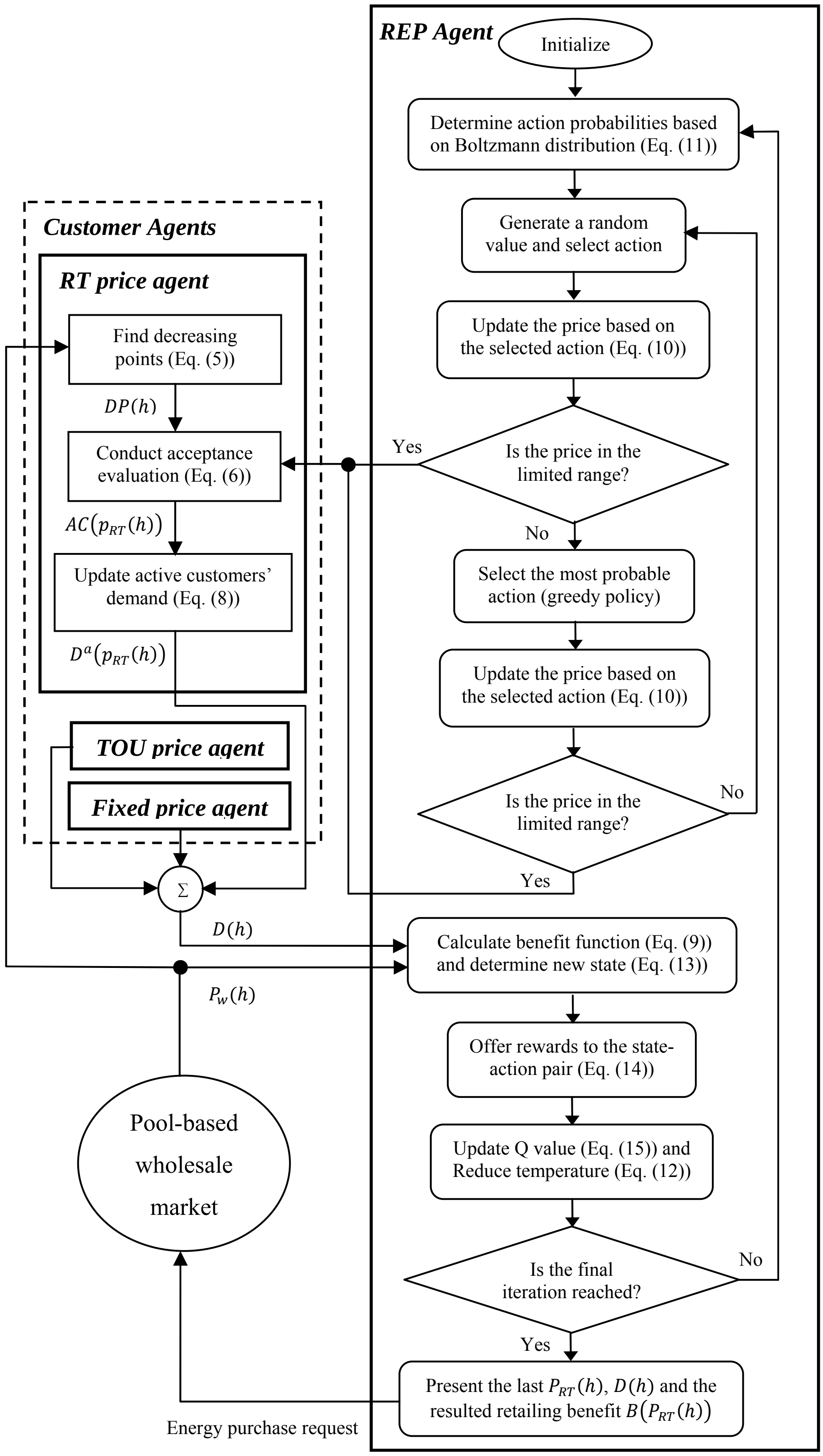

Fig. 4 The flowchart of the proposed Q-learning based real time pricing and energy procurement method.

6 Numerical Studies

In this study, the required data is extracted from the day-head market of New England, Connecticut in March to May 2010 [18]. It is considered that the REP agent supplies 30% of all Connecticut electricity demand of residential, commercial and industrial sectors which are categorized in three groups of customers, i.e. fixed, TOU and RT price agents. The target day for the retail process is 20 May 2010. The adopted parameters in the acceptance function, Q-learning approach and RTP model are provided in Table 2. It is assumed that 90% of demands in all sectors are supplied through bilateral long term contracts. Therefore, the reminder portion of it should be supplied in short term (DA and real time) markets.

It should be notified that QL approach simulates the intelligent agent's learning process which may lead to dissimilar results through several executions of the learning program. Similarly, GA is a heuristic method whose results may be unequal through different executions of the program. Consequently, each optimization process is executed 20 times and the averages of the obtained values are considered as the main results.

6.1 Evaluating the Amount of Power Needed to be Procured

The REP agent applies 24 neural networks to predict real time energy consumption of inactive customer agents based on its historical data and the latest metering data. Generally, the load profile depends significantly on the weather conditions like recorded temperature (T). The average error of forecasting temperature in the region of study [19] is applied to obtain typical forecasted temperature due to the multiplicity of factors involved in weather forecasting. The resulted percent errors of forecasting inactive customers' real time consumption is shown in Fig. 5 which demonstrates that ANNs work well in forecasting hourly demands.

The forecasted surplus consumption of inactive customer agents and the active customer agent's initial DA demand are provided in Table 3.

6.2 The performances of QL and GA in RTP for Active Customers

The REP agent learns to offer the best price through Q-learning approach. The offered price and the related benefit converge simultaneously to the final values as the learning process reaches to its limit. Fig. 6 shows the absolute relative percent difference between the resulted price in the final iteration and the best experienced price through the learning process which demonstrates the accuracy of the Q-learning approach.

In order to verify the accuracy of RTP based on Q-learning approach, its results are compared with that of GA-based RTP. Table 4 demonstrates the optimum hourly RT tariffs according to QL-based and GA-based RTP. QL and GA approaches' performance may not be judged based on their resulted price values. Their performance in determining optimum prices would be compared using the resulted retailing benefits. As it can be seen in Table 4, QL-based RTP results in %2.02 more hourly average benefit in comparison with the benefit gained by GA-based RTP.

Table 4 demonstrates better performance of the Q-learning approach. This is due to the inherent learning capability of intelligent agents which help them exploit what they already know and simultaneously explore new ways so as to make better decisions. In fact, intelligent agents exhibit obvious capabilities in modeling market players' intelligent decision making.

Table 2 Participation of customers in three pricing strategies.

Parameters	Considered Value	Equation no.
c	20	(5), (8), (9a)
m	80	(3), (6)-(9a)
σ	5	(3), (6)-(9a)
α	0.2	(15)
γ	0.95	(15)
P_{Fix}	11.5	(9a)
$P_{TOU(off-peak)}$	10.5	(9b)
$P_{TOU(peak)}$	13.5	(9b)
L	1000	(12)

Table 3 DA hourly demands of customer agents.

Hourly demands and forecasting errors	Fixed price agent	TOU price agent	RT price agent
Minimum hourly demand (MWh)	35.15	53.30	27.42
Average hourly demand (MWh)	66.90	101.45	52.20
Maximum hourly demand (MWh)	106.10	160.90	82.78
Daily demand (MWh)	1605.59	2434.84	1252.72

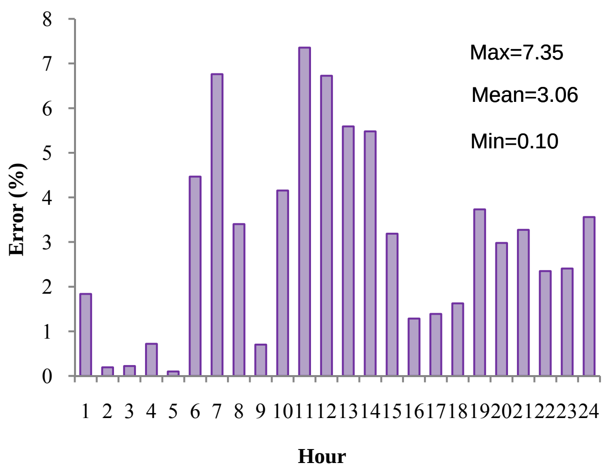

Fig. 5 Resulted errors of forecasting real time demand of inactive agents.

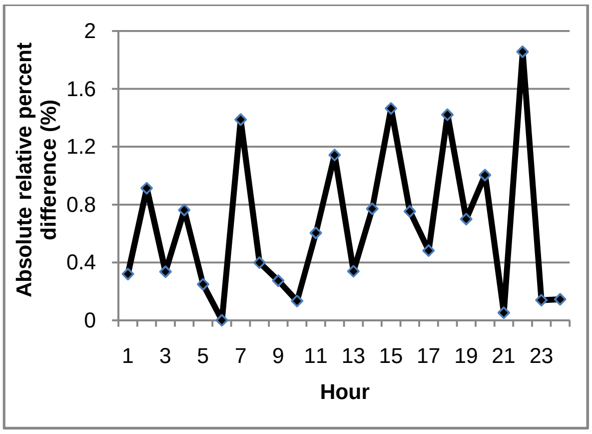

Fig. 6 The absolute relative percent difference between the resulted price in the final iteration and the best experienced price.

Table 4 The resulted RTP rates and benefits from pricing approaches based on QL and GA.

Hour	QL-based RTP rates	QL-based benefit	GA-based RTP rates	GA-based benefit
1	62.4	7369.6	62.6	7372.8
2	60.0	8734.0	60.3	8721.3
3	59.4	9984.0	59.6	9990.0
4	59.0	9951.3	59.1	9934.3
5	60.2	9268.5	60.4	9272.4
6	60.8	7131.3	61.0	7134.3
7	70.7	10507.5	71.4	10510.2
8	75.0	12431.2	74.6	12274.3
9	73.1	11019.9	73.3	11019.5
10	74.9	9605.4	74.9	9605.5
11	81.9	6815.8	82.1	6816.0
12	82.0	9454.1	82.1	9367.0
13	88.1	11163.5	87.8	11163.8
14	91.3	11765.2	89.6	11789.1
15	93.7	13459.4	93.7	13459.3
16	93.7	14691.9	93.1	14696.3
17	92.9	14138.9	93.3	14139.6
18	93.0	17957.9	92.8	17739.8
19	85.2	20732.7	85.1	20734.1
20	78.8	18653.7	78.5	18653.9
21	97.0	9676.6	96.3	9477.3
22	83.9	15303.9	84.2	15304.4
23	71.6	18988.7	71.9	18991.3
24	68.7	13774.8	69.3	13778.3
Average Hourly	77.4	12190.8	77.4	12164.4

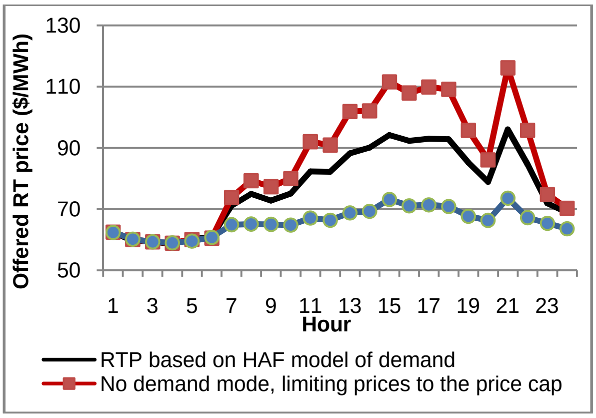

Fig. 7 The offered prices in RTP based on HAF, MSF, and the price cap.

6.3 RTP Based on HAF and MSF as Two Customer Response Models and the Price Cap

Here, the customer response is simulated using two demand models, namely MSF applied in [9] and the proposed HAF. Also, RTP is simulated considering no customer response but including a price cap for the offered prices. It means that the offered prices are not limited by the customers' reaction but they are limited to a predetermined price cap (e.g. 70% more than DA wholesale prices; $P_{RT}(h) \leq 1.7P_w(h)$).

The resulted prices of benefit function optimization using QL approach based on the proposed HAF are compared with those based on MSF and the price cap in Fig. 7. As it can be seen in Fig. 7, the hourly optimized day-ahead prices offered to the active customer agent using the proposed model are higher than the offered prices which are limited by MSF. In case of modeling customers' behavior with the previous acceptance function (MSF), it is considered that customers react to RT prices according to MSF model with no difference in peak, off-peak and valley time periods. Accordingly, the offered hourly prices are decided upon based on similar demand model through 24 hours of DA pricing. Consequently, demand modeling based on the MSF results in disappointing decreases in the optimum prices offered to the customers. However in real world, customers accept higher prices during peak hours due to the changes in hourly energy consumption and electricity procurement cost. Fig. 7 shows that based on the proposed HAF, customers respond to the hourly retail prices according to hourly energy procurement costs. Also, as it is shown in Fig. 7, the hourly optimized day-ahead prices offered to the active customers using demand response models (MSF and HAF) are less than the offered prices which are limited to the retail price cap. This is due to the sensitivity of demand to the hourly varying prices and the REP agent's competition interaction with other REPs which are modeled by the acceptance functions.

Fig. 8 compares the benefit of the REP agent in three cases of RTP based on HAF, MSF, and price caps.

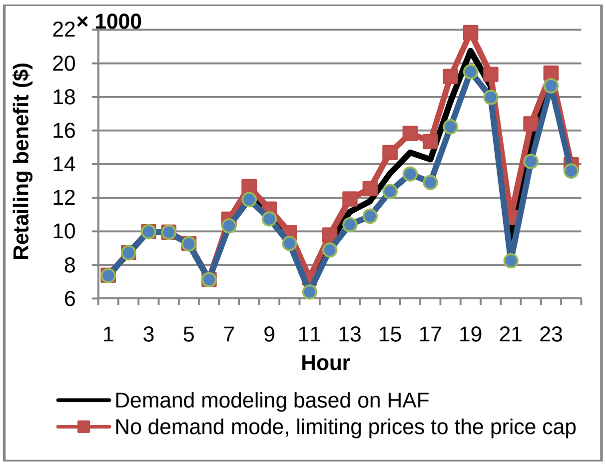

Fig. 8 Comparison of the obtained benefits in RTP based on HAF, MSF and price caps.

As it can be seen in Fig. 8 Comparison of the obtained benefits , the total benefit of the retailer based on the proposed model lies between the benefits related to other two models. This is due to the fact that in real world situation, the competition between REPs motivates the customers to reduce their demand and/or change their retailer according to the offered prices. As a result, real time pricing based on HAF model of demand leads to lower optimum prices and benefits in comparison with RTP subject to the price cap constraint. Furthermore, it can be seen that the REP agent's obtained benefit in HAF-based demand modeling is higher than the obtained benefit in MSF-based demand representation. This is due to the varying decreasing points in the proposed hourly acceptance function which result in better modeling of the rational behaviors of the customers in real retail environment.

7 Conclusion

Here, electricity procurement and sale by one of local REP agents for its customers has been modeled in an economically optimized manner and the most beneficial real time prices are determined through REP agent's learning process based on Q-learning method's principles. The better performance of the Q-learning approach compared with the performance of GA in RTP was demonstrated according to the resulted hourly benefits. The Q-learning procedure was adopted in a way to incorporate different aspects of the problem such as price caps, acceptance function and purchasing scenario. Furthermore, an hourly acceptance function has been proposed in order to model the customers' responses to the hourly retail prices. It was shown that active customers are sensitive to the offered real time prices and have the right to select other profitable REPs even a day before consumption. The acceptance function has been utilized in modeling rebutting reaction of this customer agent.

The model proposed in this paper provides the intelligence for the stakeholders as the decision makers in the retail environment. Furthermore, the proposed model features with the modularity capability which enables the user to model new entities participating in the market.

References

[1] Muller M., Sensfu X. F. and Wietschel M., "Simulation of current pricing-tendencies in the German electricity market for private consumption", *Energy Policy*, Vol. 35, No. 8, pp. 4283-4294, 2007.

[2] Lopes F., Novais A. Q. and Coelho H., "Bilateral negotiation in a multi-agent energy market", *In: De-Shuang H, Kang-Hyun J, Hong-Hee L, Hee-Jun K, Vitoantonio B (Eds) Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, pp. 655-664, 2009.

[3] Daneshfar F., Bevrani H. and Mansoori F., "Load-Frequency Control: a GA based Bayesian Networks Multi-agent System", *Iranian Journal of Electrical and Electronic Engineering, (IJEEE)*,Vol. 7, No. 2, pp. 141-148, 2011.

[4] Win O. N. and Miranda V., "Multi-energy retail market simulation with intelligent agents", *IEEE PowerTech 2005*, pp. 1-7, Russia, 2005.

[5] Penya Y. K. and Jennings N. R., "Optimal combinatorial electricity", *Web Intelligence and Agent Systems*, Vol. 6, No. 2, pp. 123-135, 2008.

[6] Bompard E. F., Napoli R., Abrate G. and Wan B., "Multi-agent models for consumer choice and retailer strategies in the competitive electricity market", *International Journal of Emerging Electric Power Systems*, Vol. 8, No. 2, 2007.

[7] Zare K., Parsa Moghaddam M. and Sheikh-El-Eslami M. K., "Electricity procurement for large consumers based on Information Gap Decision Theory", *Energy Policy*, Vol. 38, No. 1, pp. 234-242, 2010.

[8] Rabiee A., Shayanfar H. and Amjady N., "A New Framework for Reactive Power Market Considering Power System Security", *Iranian Journal of Electrical and Electronic Engineering (IJEEE)*, Vol. 5, No. 3, pp. 196-204, 2009.

[9] Hatami A. R., Seifi H. and Sheikh-El-Eslami M. K.," Optimal selling price and energy procurement strategies for a retailer in an electricity market", *Electric Power System Research*, Vol. 79, No. 1, pp. 246-254, 2009.

[10] Barforoushi T., Parsa Moghaddam M., Javidi M. H. and Sheik-El-Eslami M. K., "A New Model Considering Uncertainties for Power Market Simulation in Medium-Term Generation Planning", *Iranian Journal of Electrical and Electronic Engineering (IJEEE)*, Vol. 2, No. 2, pp. 71-81, 2006.

[11] Alami H. A., Parsa Moghaddam M. and Yousefi G. R., "Demand response modeling considering Interruptible/Curtailable loads and capacity

market programs”, *Applied Energy*, Vol. 87, No. 1, pp. 243-250, 2010.

[12] Yusta J. M. , Khodr H. M. and Urdaneta A. J.,” Optimal pricing of default customers in electrical distribution systems: Effect behavior performance of demand response models”, *Electric Power System Research*, Vol. 77, No. 5/6, pp. 548-558, 2007.

[13] Aalami H. A., Parsa Moghaddam M. and Yousefi G. R., “Modeling and prioritizing demand response programs in power markets”, *Electric Power System Research*, Vol. 80, N0. 4, pp. 426-435, 2010.

[14] Herterand K. and Wayland S., “Residential response to critical-peak pricing of electricity: California evidence”, *Energy*, Vol. 35, N0. 4, pp. 1561-1567, 2010.

[15] Gabriel S. A., Conejo A. J., Plazas M. A. and Balakrishnan S., “Optimal price and quantity determination for retail electric power contracts”, *IEEE Trans. Power System*, Vol. 21, No. 1, pp. 180-187, 2006.

[16] Mahmoudi-Kohan N., Parsa Moghaddam M. and Sheikh-El-Eslami M. K., “An annual framework for clustering-based pricing for an electricity retailer”, *Electric Power System Research*, Vol. 80, No. 9, pp. 1042-1048, 2010.

[17] Kotler P., *Armstrong G Principles of marketing*, Prentice Hall, 2005.

[18] Independent System Operator (ISO) New England. Hourly zonal information, available at: http://www.iso-ne.com

[19] 5-day weather forecast for Windsor Locks, Connecticut, available at: http://www.forecastadvisor.com

Shaghayegh Yousefi received her B.Sc. degree in electrical engineering from the electrical engineering department of the University of Tehran, in 2003 and the M.Sc. degree from the Tarbiat Modares University, Tehran, Iran, in 2005. She is currently a PHD student in the faculty of electrical and computer engineering, Tarbiat Modares University, Tehran, Iran. Her areas of interest include smart power grids and multi-agent systems.

Mohsen Parsa Moghaddam was born in Iran, 1956. He received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1980, the M.Sc. degree from the Toyohashi University of Technology, Japan, in 1985, and the Ph.D. degree from Tohoku University, Japan, in 1988. Currently, he is a Professor in the Faculty of the Electrical and Computer Engineering, Tarbiat Modares University (TMU), Tehran, Iran. His research interests include power system planning & control, energy management, and smart grid.

Vahid Johari Majd received his B.Sc. degree in 1989 from the electrical engineering department of the University of Tehran, Iran. He then received his M.Sc. and Ph.D. degrees in the area of Control Theory from the electrical engineering department of the University of Pittsburgh, PA, U.S.A. in 1991 and 1995, respectively. He is currently an associate professor in the control system department of Tarbiat Modares University, Tehran, Iran, and is the director of intelligent control systems laboratory. His areas of interest include intelligent identification and control, multi-agent learning, fuzzy control, robust control, nonlinear control, fractional order control, chaotic systems, and systems biology.