International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:6, No:7, 2012

Q-Learning with Eligibility Traces to Solve Non-Convex Economic Dispatch Problems

Mohammed I. Abouheaf, Sofie Haesaert, Wei-Jen Lee, Frank L. Lewis

***Abstract*—**Economic Dispatch is one of the most important power system management tools. It is used to allocate an amount of power generation to the generating units to meet the load demand. The Economic Dispatch problem is a large scale nonlinear constrained optimization problem. In general, heuristic optimization techniques are used to solve non-convex Economic Dispatch problem. In this paper, ideas from Reinforcement Learning are proposed to solve the non-convex Economic Dispatch problem. Q-Learning is a reinforcement learning techniques where each generating unit learn the optimal schedule of the generated power that minimizes the generation cost function. The eligibility traces are used to speed up the Q-Learning process. Q-Learning with eligibility traces is used to solve Economic Dispatch problems with valve point loading effect, multiple fuel options, and power transmission losses.

***Keywords*—**Economic Dispatch, Non-Convex Cost Functions, Valve Point Loading Effect, Q-Learning, Eligibility Traces.

I. Introduction

THE operation of the power system is tightly controlled to achieve the efficient use of its capabilities [1]. The operation cost of the generating units highly depends on the fuel cost. The Economic Dispatch (ED) is a modern power system energy management tool. It results in the best economical use of the generating units and fuel sources [2].

The high nonlinearity of the power system imposes mathematical complexities in formulating the generation cost models necessary to solve the Economic Dispatch problem [1]. The sources of the mathematical complexities are due to the design specifications and operation constraints of the generating units such as the spinning reserve, transmission losses, prohibited operation zones, ramp rate limit, valve point loading effect, and multiple fuel options [1]. The spinning reserve determines how the generating units are robust to the unexpected outages or incorrect load allocation among the generating units [3]. The Prohibited zones are caused by faults in the machines its self or the associated auxiliaries [4]. Restrictions in the power generation define the ramp rate limits constraints [3]. In addition, some turbines use multiple valves that are opened sequentially to satisfy the load requirement, which adds more non-convexity to the generation cost function [3]. Some generating units use multiple fuel types for different operation regions [5].

The conventional methods used to solve Economic Dispatch problem include Newton Raphson, gradient techniques, lambda iteration method, the base point and participation factors method [1], interior point algorithm, linear programming, dynamic programming and dual quadratic programming [6]-[8] where the generation cost functions are assumed to be monotonically increasing piece-wise linear functions. Heuristic optimization techniques are used to find the optimal solution for the non-convex Economic Dispatch problem. These techniques include Evolutionary programming (EP) [9], Genetic Algorithm (GA) [4], Differential Evolution [10], Particle Swarm Optimization (PSO) [11], Simulated annealing (SA) [12], Tabu Search [13], Gravitational Search Algorithm (GSA) [2], and Biogeography method [14]. These Heuristic algorithms don't always guarantee the global best solution.

In this paper, ideas from Reinforcement Learning (RL) are used to solve the non-convex Economic Dispatch problem. Reinforcement Learning is an area of machine learning, used to solve multi-stage decision making problems. It is concerned with how an agent will pick its actions in a dynamic environment to transit to new states in such a way that the optimization of the objective function can be achieved [15]-[20].

This paper is organized as follows. In Section II, the classical Economic Dispatch problem is introduced. In Section III, Q-Learning and eligibility traces are introduced. In Section IV, an algorithm based on Q-Learning with eligibility traces is developed to solve non-convex Economic Dispatch. In Section V, simulation is performed using the developed algorithm to solve the Economic Dispatch problem with valve point loading effect, multiple fuel options, and transmission losses.

II. Formulation of the Economic Dispatch Problem

In this section the classical Economic Dispatch problem is formulated using Lagrange dynamics [1]. The main operation constraints related to the generating units are mentioned. Furthermore, the different generation cost models are introduced.

A. Economic Dispatch Problem

Lagrange dynamics is used to formulate and solve the Economic Dispatch problem. The objective of the optimization problem is to minimize the fuel generation cost, so that

Mohammed I. Abouheaf is with the Automation and Robotics Research Institute, The University of Texas at Arlington, 7300 Jack Newell Blvd. S., Ft. Worth, TX 76118 USA (Phone: +1 817-272-5955; Fax:+1 817-272-5952; e-mail: abouheaf@uta.edu).
Sofie Haesaert is with Delft University of Technology, Nederlands (e-mail: s.haesaert@student.tudelft.nl).
Wei-Jen Lee is with the Energy Systems Research Center, The University of Texas, Arlington, TX 76013 USA (e-mail:wlee@uta.edu).
Frank L. Lewis is with the Automation and Robotics Research Institute, The University of Texas at Arlington, 7300 Jack Newell Blvd. S., Ft. Worth, TX 76118 USA (e-mail: lewis@uta.edu).

$$Minimize \quad F_T = \sum_{i=1}^{ng} F_i(P_i), \quad \forall i \tag{1}$$

where F_T is the total fuel generation cost and it is given by $F_T = F_1 + F_2 + F_3 + ... + F_{n_g}$, F_i is the fuel generation cost of each unit *i*, P_i is the power generated by each unit *i*, and n_g is the number of generating units.

The generation cost function F_i is approximated by the quadratic function [1], so that

$$F_i(P_i) = a_i + b_i P_i + c_i P_i^2 \tag{2}$$

where a_i, b_i, and c_i are the fuel cost coefficients of the generation unit *i*.

Equation (2) states the basic generation cost model. The Lagrange operator is given so that

$$L = F_T + \sum_{i=1}^{N_c} \lambda_i \varphi_i \tag{3}$$

where N_c is the number of constraints, λ_i is the Lagrange multiplier associated with each constraint φ_i.

The Lagrange operator *L* is minimized with respect to the generated power, while the constraints $\varphi_i, \forall i$ are satisfied [1].

B. Operation Constraints

The generating units' constraints are classified into two types. The first is related to the design and operation specifications of the generating units such as the generation capacity, line maximum power flow, generation ramp limits, prohibited operation zones constraints, and spinning reserve. The second is related to an upper level of operation control such as unit commitment and other operation plans like maintenance. Here, only constrains related to the work are considered.

1) Generation-Demand Equality Constraints

The Generation–Demand equality constraint, states that the sum of the generated power is equal to the total active load demand plus the transmission losses so that

$$\sum_{i=1}^{ng} (P_i) = P_D + P_{Losses}, \tag{4}$$

where P_D is the total active load demand, P_{Losses} is the transmission losses. The transmission losses is given in terms of Kron's loss formula [5] so that

$$P_{Losses} = \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} \left(P_i B_{ij} P_j \right) + \sum_{i=1}^{n_g} \left(B_{oi} P_i \right) + B_{oo} \tag{5}$$

where B_{ij}, B_{oi}, and B_{oo} are the transmission network power losses coefficients. The B-loss coefficients represent the transmission line and the corona losses [5].

2) Generation Capacity

Each generating unit has maximum and minimum generation capacities so that

$$P_i^{\min} \le P_i \le P_i^{\max}, \qquad \forall i \tag{6}$$

where $P_i^{\min}$ and $P_i^{\max}$ are the designed minimum and maximum generated power capacities of each unit *i*.

3) Spinning Reserve Constraints

During the power system operation, the generating units are not working on the maximum designed capacity, instead those units keep about 5-10% of their capacity unused [21]. This operation enhances the security of the power system in the case of emergencies. Here, the spinning reserve constraints are given only for the generating units without prohibition zones [21] so that

$$SR_i = \min\left\{ (P_{i(\max)} - P_i), SR_{i(\max)} \right\}, \ \forall i \ without \ POZ \tag{7}$$

$$SR = \sum_{i=1}^{n_g} SR_i \tag{8}$$

where SR_i is the spinning reserve of unit *i* (MW), $SR_{i(\max)}$ is the maximum spinning reserve of unit *i*, SR is the total spinning reserve given by the generating units that do not have any Prohibited Operating Zones (POZ).

C. Practical Generation Cost Functions

The simplified generation cost function (2) does not include the valve point loading effect and the multiple fuel types' effects. Accurate generation cost models are given as follows.

1) Economic Dispatch with Valve Point Loading Effect

The admission valves operate in a sequential manner in some turbines. This sequential operation causes ripples or non-differentiable points in the generation cost models [22]. This effect is modeled by a sinusoidal function so that, the generation cost function is given by

$$F_i(P_i) = a_i + b_i P_i + c_i P_i^2 + \left| e_i \times \sin(f_i \times \sin(P_{i(\min)} - P_i)) \right| \tag{9}$$

where a_i, b_i, c_i, e_i, and f_i are the fuel cost coefficients for each unit *i* and $P_{i(\min)}$ is the minimum generated power by each unit *i* with valve point loading effect.

2) Economic Dispatch with Multiple Fuel Options

The generating units can use multiple fuel options for different regions in the operation range. This adds more non-convexity to the generation cost function so that

$$F_i(P_i)=\begin{cases} a_{i1}+b_{i1}P_i+c_{i1}P_i^2, & P_{i(\min)}\le P_i\le P_{i1} \\ a_{i2}+b_{i2}P_i+c_{i2}P_i^2, & P_{i1}\le P_i\le P_{i12} \\ \vdots & \vdots \\ a_{ik}+b_{ik}P_i+c_{ik}P_i^2, & P_{i(k-1)}\le P_i\le P_{i(\max)} \end{cases} \tag{10}$$

where a_{ik}, b_{ik}, and c_{ik} are the generation cost coefficients of each unit *i* using fuel type *k*.

3) Economic Dispatch with Valve Point Loading Effect and Multiple Fuel Options

The generation cost function due to valve point loading effect and multiple fuel options is denoted as "Hybrid cost function" [22]. The hybrid cost function models result from combining the generation cost models (9) and (10) so that

$$F_i(P_i)=\begin{cases} a_{i1}+b_{i1}P_i+c_{i1}P_i^2+\left|e_{i1}\times\sin(f_{i1}\times\sin(P_{i(\min)}-P_i))\right|, \\ \qquad (P_{i(\min)}\le P_i\le P_{i1}) \\ a_{i2}+b_{i2}P_i+c_{i2}P_i^2+\left|e_{i2}\times\sin(f_{i2}\times\sin(P_{i1}-P_i))\right|, \\ \qquad (P_{i1}\le P_i\le P_{i12}) \\ \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\ a_{ik}+b_{ik}P_i+c_{ik}P_i^2+\left|e_{ik}\times\sin(f_{ik}\times\sin(P_{i(k-1)}-P_i))\right|, \\ \qquad (P_{i(k-1)}\le P_i\le P_{i(\max)}) \end{cases} \tag{11}$$

where a_{ik}, b_{ik}, c_{ik}, e_{ik}, and f_{ik} are the fuel cost coefficients for each unit *i* and fuel type *k*.

III. Reinforcement Learning with Eligibility Traces

In this section, the Reinforcement Learning (RL) is used to solve the Economic Dispatch problem with valve point loading effect, multiple fuel options, and transmission losses. Ideas from Reinforcement Learning, Markov Decision process, Q-Learning, and eligibility traces are introduced [16].

The Reinforcement Learning (RL) algorithm developed herein learns the optimal power distribution for the Economic Dispatch problem by interacting with the environment i.e. choosing the proper generating values to minimize the generation cost objective functions [23].

A. Markov Decision Process

Reinforcement Learning requires a mapping of the continuous Economic Dispatch problem structure to a discrete problem structure similar to the Markov Decision Process (MDP) [20].

The normal Markov Decision Process (MDP) is defined by the tuple $M\langle X,U,f,\rho\rangle$ where X is the discrete set of all possible states, U is the discrete set of all possible actions. $f: X\times U\to X$ is the state transition function, and $\rho: X\times P\to R^1$ is the penalty function. The actions are chosen based on a policy $\pi: X\mapsto U$. This policy minimizes the sum of future costs, this sum is stored in a value function.

B. Q-Learning

Q-Learning is a reinforcement learning technique. The goal of each agent (generating unit) is to learn a policy (scheduling the generated power) that minimizes the penalty function (generation cost function) (1). One way to learn the optimal policy (optimal generation schedule) is by using Q-Learning with the sum of future costs to be defined by the Q-function $Q: X\times U\to R^1$. The Q-function gives the expected cost for a given state-action pair under a given policy π so that

$$Q(x,u)=\rho(x,u)+\min_{u'}Q(f(x,u),u') \tag{12}$$

where $\rho(x,u)$ is the penalty function.

The Q-function is iterated by correcting the old value with the penalty so that

$$\begin{aligned} &Q^{q+1}(x_k,a_k)=Q^q(x_k,a_k)+ \\ &\alpha(\rho(x_k,a_k)+\gamma\min_{\overline{a}\in A}(Q(x_{k+1},\overline{a}))-Q^q(x_k,a_k)) \end{aligned} \tag{13}$$

where $\rho(x_k,a_k)$ is the penalty function (generation cost function), α is the learning coefficient, *q* is the number of the iterations, *k* is the number of the state, γ is the discount factor, and *A* is the set of all possible actions.

The balance between the exploration and exploitation is important in the Q-Learning. Moreover, the proper selection of the action affects the performance of both the learning and evaluation of the agent's policy [16]. The ε greedy (near-greedy) method is an effective strategy of choosing the best actions during Q-Learning. It acts very well in environments with noisier cost functions. The ε *-greedy* Q-Learning selects the action with the lowest expected cost with probability $1-\varepsilon$ and selects a random action from the feasible action set with probability $\varepsilon\in[0,1]$ [16].

The $Q(\overline{\lambda})$ learning with eligibility traces is used to speed up the Q-Learning process. As per Sutton and Barto, the eligibility trace $\overline{\lambda}$ temporarily memorizes the parameters associated with an event to be eligible for learning changes in the Q-Learning process [16]. The state–action pairs are backed together and memorized as long as the greedy policy is followed.

IV. $Q(\overline{\lambda})$ Learning with Eligibility Traces Algorithm

In this section, an algorithm based on Q-Learning with eligibility traces is developed. Algorithm 1 explains how the Q-Learning algorithm with eligibility traces will learn the optimal generated powers for any applicable active load demand. Next, Algorithm 2 is used to extract the optimal actions (optimal generated power instances) for specified active load demands, taking into consideration the transmission losses. The Markov Decision Process implies that the different stages are seen as the different generation units. The state x_k is defined as the residual power demand

and k is the generator number. The power demand and the action spaces are discretized, whenever the generated power space is originally continuous. The discretization step will be an important factor that will impact the accuracy of the results. The action space $U(x_k)$ is the feasible generated power choices for the state x_k. The penalty function is given in terms of the generation cost functions ((9)-(11)). The Q-Learning process Algorithm is given as follows

Algorithm 1: Q-Learning with Elgibility Traces (Learning Process)

- Identify the minimum and the maximum possible generated power for all the generation units n_g.
- Determine the power demand limits, the demand should not be greater than the summation of maximum generated power or not less than the summation of minimum generated power.

Assume that every generating unit, generates at least its minimum power so the maximum amount of power that needs to be distributed over the generation units is given by

$$P_{D\max} = \sum_{\forall ng} (P_{\max} - P_{\min})$$

- Initialize the Q-function, the total number of trials (iterations), and the exploration rate ε.

while $\left(t_r \le trial_{\max}\right)$ Do {

1. Generate random power demand instance (P_D) picked from the uniform distribution over [0, $P_{D\max}$].
2. Define the first state so that $x_1 = (1,\ P_D)$.
3. Determine optimal action for the first generation unit by

3.1. Identify the feasible discrete action space $P_1 \in U(x_1)$ so that

$$(x_1 - \sum_{j=2}^{n_g} (P_{\max})_j + \sum_{j=2}^{n_g} (P_{\min})_j) \le P_1 \le x_1,$$
$$0 \le P_1 \le \left((P_{\max})_1 - (P_{\min})_1\right)$$

3.2. Retrieve the optimal action $P_1^* = \arg\min_{P_1^{'} \in U(x_1)} Q(x_1, P_1^{'})$ For the next states steps

For $k = 1,.....,n_g - 2$ Do {

4. Apply $\bar{\varepsilon}$-greedy action

If $\bar{\varepsilon} < (1-\varepsilon)$ do { $P_k = P_k^*$

Otherwise $P_k = rand\{U(x_k)\}$

Storing the k index to be used for the eligibility trace $k_r := k$ }

5. The remaining load to be distributed to the next stages or states. State transfer: $x_{k+1} = x_k + \langle 1, -P_k \rangle$.
6. Determine optimal action for $(k+1)^{\text{th}}$ generator by

6.1. Identify the feasible action space $P_{k+1} \in U(x_{k+1})$ so that

$$(x_{k+1} - \sum_{j=k+2}^{n_g} (P_{\max})_j + \sum_{j=k+2}^{n_g} (P_{\min})_j) \le P_{k+1} \le x_{k+1},$$
$$0 \le P_{k+1} \le \left((P_{\max})_{k+1} - (P_{\min})_{k+1}\right)$$

6.2. Retrieve the optimal action P_{k+1}^* from the feasible space such that $P_{k+1}^* = \arg\min_{P_{k+1}^{'} \in P_{k+1,\, feasible}} Q(x_{k+1}, P_{k+1}^{'})$.

7. Update Q-function including trace information

Define the error function (Δ_k) so that

$$\Delta_k = F(P_k) + Q^q(x_{k+1}, P_{k+1}^*) - Q^q(x_k, P_k)$$

For (x_l, p_l) with $l \in [k_r, ..., k]$

$$Q^{q+1}(x_l, p_l) = Q^q(x_l, p_l) + \alpha \bar{\lambda}^{k-1} [\Delta_k]$$

where $\bar{\lambda} = 1 - \exp(-t_r / trial_{\max})$

End

End}

8. Repeat steps (3) and (4) for $k = n_g - 1$
9. The feasible action space of last generator is given by

$$P_{ng} = P_D - \sum_{k=1}^{n_g - 1} P_k$$

10. Update Q-function

Calculate the error function for the last stage (generation unit).

$$\Delta_k = F(P_{n_g - 1}) + F(P_{n_g}) - Q^q(x_{n_g - 1}, P_{n_g - 1})$$

For $\left(x_l, p_l\right)$ with $l \in [k_r, ..., k]$

$$Q^{q+1}(x_l, p_l) = Q^q(x_l, p_l) + \alpha \bar{\lambda}^{k-1} [\Delta_k]$$

End}

Algorithm 2 extracts the optimal power distributions for a given active load demand, taking into consideration the transmission losses.

Algorithm 2: Extracting the optimal power distribution considering the transmission losses.

- Define the required tolerance μ (Convergence error).
- Identify possible state-action pairs for all stages (results from Algorithm 1 (learning process)).
- Initialize the error coefficient δ, which describes the difference between the power losses for the successive iterations.
- Initialize $P_{Expected\ Losses}$

For a given active load demand P_D do the following:

while $\delta > \mu$ *Do* {

1. The modified load demand is $\bar{P}_D = P_D + P_{Expected\ Losses}$.
2. Use the results from Algorithm 1 to obtain the optimal generated power vector $\left(P_T\right)$ knowing $\bar{P}_D$.
3. Calculate the expected transmission losses $P_{Losses}\left(P_T\right)$ using (5).
4. Update error δ and the expected losses $P_{Expected\ Losses}$ so that $\delta = P_{Expected\ Losses} - P_{Losses}$, $P_{Expected\ Losses} = P_{Losses}$ }.

Initially the computational effort is done in the learning process. Once the learning process is done, it is easy to retrieve the optimal power distribution for the generating units for any active load demand [20]. To simplify the simulation of the Q learning with eligibility traces, the learning coefficient is picked so that $\alpha = 1$, where it is multiplied by the exponentially decreasing trace $\bar{\lambda}^k$, and the discount factor is picked so that $\gamma = 1$.

V. Q-Learning: Case Studies and Numerical Simulation

The advantages of the proposed algorithm to solve the Economic Dispatch problem are verified in this section. The Q-Learning with eligibility traces is compared to other Heuristic optimization techniques. Three study cases are considered for the simulation purposes. In case 1, the Q-Learning with eligibility traces is used to solve Economic Dispatch problem for 6 generating units with valve point loading effect and transmission losses. In case 2, the Q-Learning with eligibility traces is used to solve Economic Dispatch problem for 10 generating units with multiple fuel options. In case 3, Q-Learning with eligibility traces is used to solve Economic Dispatch for 15 generating units considering the transmission losses.

A. Case Study 1:

In this case, Q-Learning with eligibility traces algorithm results is compared to other published results for 6 generating units. The fuel cost coefficients and generation capacities of the 6 generating units with valve point loading effect are given in Table I. The simulation parameters (discrete step=7 MW, $trail_{\max} = 10^5$, $\varepsilon = 0.1$, and $\mu = 0.01$).

TABLE I
CASE STUDY 1: GENERATION CAPACITIES AND COST COEFFICIENTS OF SIX THERMAL GENERATION UNITS

Unit	a($)	b($/MW)	c($/MW^2)	E	f	P_{min}	P_{max}
1	240	7	0.007	300	0.031	100	500
2	200	10	0.0095	200	0.042	50	200
3	220	8.5	0.009	150	0.063	80	300
4	200	11	0.009	150	0.063	50	150
5	220	10.5	0.008	150	0.063	50	200
6	190	12	0.0075	150	0.063	50	120

The transmission power losses are expressed in terms of Kron's loss formula. The losses coefficients B_{ij}, B_{oi}, and B_{oo} are given as follows

$$B_{ij} = \begin{bmatrix} 0.0017 & 0.0012 & 0.0007 & -0.0001 & -0.0005 & -0.0002 \\ 0.0012 & 0.0014 & 0.0009 & 0.0001 & -0.0006 & -0.0001 \\ 0.0007 & 0.0009 & 0.0031 & 0 & -0.001 & -0.0006 \\ -0.0001 & 0.0001 & 0 & 0.0024 & -0.0006 & -0.0008 \\ -0.0005 & -0.0006 & -0.001 & -0.0006 & 0.0129 & -0.0002 \\ -0.0002 & -0.0001 & -0.0006 & -0.0008 & -0.0002 & 0.015 \end{bmatrix}$$

Bo=0.001*[-0.3908 -0.1297 0.7047 0.0591 0.2161 -0.6635]. Boo=0.0056.

The optimal generated power for different methods are given in Table II for active load demand (PD=1263 MW). The Q-Learning with eligibility traces achieved the lowest fuel generation cost (15452 $/h) compared to GA [24], RGA [25], and PSO [14] as shown in Table II. For active load demand (PD=1262 MW) Q-Learning with eligibility traces achieved the lowest fuel generation cost (15439 $/h) compared to BBO [26], and BGA [14] and it was the same as IWD [27] as shown in Table III. Moreover, results for Q-Learning with eligibility traces are given in Table IV for active load demands (1080 MW, 1100 MW, 1220 MW, and 1240 MW).

TABLE II
COMPARISON BETWEEN Q-LEARNING WITH ELIGIBILITY TRACES AND OTHER METHODS WITH TRANSMISSION LOSSES FOR ACTIVE LOAD DEMAND (PD=1263 MW) (WITHOUT VALVE EFFECT LOADING POINT)

Unit	GA [24]	RGA[25]	PSO [14]	Q-Learning
P1(MW)	474.807	420.2342	432.9639	448.9480
P2(MW)	178.636	199.4412	170.5198	173.5954
P3(MW)	262.209	263.7234	261.9009	266.2876
P4(MW)	134.283	120.0030	116.9111	127.1212
P5(MW)	151.904	167.2319	190.4102	174.3471
P6(MW)	74.181	105.1250	103.4931	85.9702
Losses	13.022	13.2627	13.142	13.274
T. G. P.	1276.03	1275.8	1276.2	1276.3
T. G. C.($/h)	15459	15461	15458.56	15452

TABLE III
COMPARISON BETWEEN Q-LEARNING WITH ELIGIBILITY TRACES AND OTHER METHODS WITH TRANSMISSION LOSSES FOR ACTIVE LOAD DEMAND (PD=1262 MW) (WITHOUT VALVE EFFECT LOADING POINT)

Unit	BBO [26]	BGA[14]	IWD[27]	Q-Learning
P1(MW)	447.3997	447.0877	450.13	448.9601
P2(MW)	173.2392	173.1887	173.62	173.4225
P3(MW)	263.3163	263.9242	260.61	266.0719
P4(MW)	138.0006	138.0607	139.49	126.9164
P5(MW)	165.4104	165.5524	159.7	174.1355
P6(MW)	87.07979	86.6289	90.51	85.7446
Losses	12.44	12.4465	12.05	13.2554
T. G. P.	1274.44	1274.443	1274.05	1275.3
T. G. C. ($/h)	15443	15443	15439	15439

TABLE IV
Q-LEARNING WITH ELIGIBILITY TRACES FOR DIFFERENT ACTIVE LOAD DEMANDS WITH VALVE POINT LOADING EFFECT AND TRANSMISSION LOSSES

Unit	1080	1100	1220	1240
P1	402.8731	405.0605	406.3627	404.3976
P2	121.3979	125.5647	197.3124	125.0499
P3	276.2484	281.8378	281.7767	279.2372
P4	96.3890	99.1161	99.0728	147
P5	97.7152	148	148.9676	198.4426
P6	95.4898	51.1164	99.2119	98.8735
Losses	10.1128	10.7016	12.6945	13.0051
T. G. P.	1090.1	1110.7	1232.7	1253
T. G. C. ($/h)	13239	13383	14991	15244

B. Case Study 2

In this case, Q-learning with eligibility traces algorithm results is compared to other published results for 10 generating units. The fuel cost coefficients and generation capacities of the 10 generating units with valve point loading effect and multiple fuel options are given in Table V and Table VI. The simulation parameters (discrete step=7 MW, $trail_{\max} = 3*10^5$, $\varepsilon = 0.1$, and $\mu = 0.01$).

TABLE V
CASE STUDY 2: FUEL COST COEFFICIENTS, VALVE POINT LOADING EFFECT, AND FUEL TYPES FOR 10 GENERATING UNITS

Unit	Fuel	a	b	c	e	f
1	1	26.97	-0.3975	0.002176	0.027	-4
	2	21.13	-0.3059	0.001861	0.0211	-3.1
2	1	1.865	-0.03988	0.001138	0.0019	-0.4
	2	13.65	-0.198	0.00162	0.0137	-2
	3	118.4	-1.269	0.004194	0.1184	-13
3	1	39.79	-0.3116	0.001457	0.0398	-3.1
	2	-2.875	0.03389	0.0008035	-0.003	0.3
	3	-59.14	0.4864	0.00001176	-0.059	4.9
4	1	1.983	-0.03114	0.001049	0.002	-0.3
	2	52.85	-0.6348	0.002758	0.0529	-6.3
	3	266.8	-2.338	0.005935	0.2668	-23
5	1	13.92	-0.08733	0.001066	0.0139	-0.9
	2	99.76	-0.5206	0.001597	0.0998	-5.2
	3	-53.99	0.4462	0.0001498	-0.054	4.5
6	1	1.983	-0.03114	0.001049	0.002	-0.3
	2	52.85	-0.6348	0.002758	0.0529	-6.3
	3	266.8	-2.338	0.005935	0.2668	-23
7	1	18.93	-0.1325	0.001107	0.0189	-1.3
	2	43.77	-0.2267	0.001165	0.0438	-2.3
	3	-43.35	0.3559	0.0002454	-0.043	3.6
8	1	1.983	-0.03114	0.001049	0.002	-0.3
	2	52.85	-0.6348	0.002758	0.0529	-6.3
	3	266.8	-2.338	0.005935	0.2668	-23
9	1	14.23	-0.01817	0.0006121	0.0142	-0.2
	2	88.53	-0.5675	0.001554	0.0885	-5.7
	3	14.23	-0.01817	0.0006121	0.0142	-0.2
10	1	13.97	-0.09938	0.001102	0.014	-1
	2	46.71	-0.2024	0.001137	0.0467	-2
	3	-61.13	0.5084	0.00004164	-0.061	5.1

TABLE VI
CASE STUDY 2: FUEL OPTIONS AND GENERATION UNITS' CAPACITIES

Unit	Pmin	Fuel	P_1		P_2	Fuel	Pmax
1	100	Fuel 1	196			Fuel 2	250
2	50		114	Fuel 2	157	Fuel 3	230
3	200		332		388		500
4	99		138		200		265
5	190		338		407		490
6	85		138		200		265
7	200		331		391		500
8	99		138		200		265
9	130		213		370		440
10	200		362		407		490

For active load demand (PD=2700 MW), the Q-Learning with eligibility traces algorithm achieved the lowest fuel generation cost (624.3116 $/h) compared to HM[28], HNN[29], AHNN[30], EP [31], CGA-MU[22], IGA-MU [22], DE[32], RGA[4], PSO [32], and GA [33] as shown in Table VII. The optimal generated powers and the respective fuel types for active load demand (PD=2700 MW) are given in Table VIII.

TABLE VII
COMPARISON BETWEEN Q-LEARNING WITH ELIGIBILITY TRACES AND OTHER METHODS FOR ACTIVE LOAD DEMAND (PD=2700 MW)

Method	Cost, $/h	Method	Cost, $/h
HM [28]	625.18	DE [32]	624.5146
HNN[29]	626.12	RGA [4]	624.5081
AHNN [30]	626.24	PSO [32]	624.5074
EP [31]	626.26	GA [33]	624.5050
CGA-MU[22]	624.7193	Q-Learning	624.3116
IGA-MU [22]	624.5178		

TABLE VIII
OPTIMAL GENERATED POWER AND RESPECTIVE FUEL OPTIONS BY Q-LEARNING WITH ELIGIBILITY TRACES FOR (PD=2700 MW)

Unit	P (MW)	Fuel	Unit	P (MW)	Fuel
1	219.0810	2	6	240.2003	3
2	211.4611	3	7	288.2319	1
3	282.1781	1	8	239.3739	3
4	239.8034	3	9	426.4890	3
5	279.5825	1	10	273.5989	1
Total Generation (MW)			2700		
Total Cost ($/h)			624.3116		

C. Case study 3:

In this case, Q-learning with eligibility traces are compared to other published results for 15 generating thermal units, whose fuel cost characteristics and generation capacities are given in Table IX. Moreover, the power system transmission losses are considered. The B loss formula is used to express the transmission losses and the losses coefficients are given in the Appendix. The simulation parameters (discrete step=5 MW, $trail_{\max} = 10^5$, $\varepsilon = 0.1$, and $\mu = 0.01$).

TABLE IX
CASE STUDY 3: FUEL COST COEFFICIENTS AND GENERATION CAPACITIES FOR 15 GENERATING UNITS

Unit	a	b	c	Pmin	Pmax
1	671	10.1000	0.000299	150.0000	455.0000
2	574	10.2000	0.000183	150.0000	455.0000
3	374	8.8000	0.001126	20.0000	130.0000
4	374	8.8000	0.001126	20.0000	130.0000
5	461	10.4000	0.000205	150.0000	470.0000
6	630	10.1000	0.000301	135.0000	460.0000
7	548	9.8000	0.000364	135.0000	465.0000
8	227	11.2000	0.000338	60.0000	300.0000
9	173	11.2000	0.000807	25.0000	162.0000
10	175	10.7000	0.001203	25.0000	160.0000
11	186	10.2000	0.003586	20.0000	80.0000
12	230	9.9000	0.005513	20.0000	80.0000
13	225	13.1000	0.000371	25.0000	85.0000
14	309	12.1000	0.001929	15.0000	55.0000
15	323	12.4000	0.004447	15.0000	55.0000

Q-Learning algorithm achieved the lowest fuel generation cost (32676$/h) for active load demand (PD=2630 MW) compared to PSO [34], GA[34], SPSO [11], PC_PSO [11],

SOH-PSO [11], [34], MTS [33], SA [35], SCA [35], APSO [36], CPSO [34], BF [34], MDE [34], TSA [33], and DSPSO-TSA [33] as shown in

TABLE XI shows the optimal generated power calculated by Q-Learning with eligibility traces for active load demand (PD=2630 MW).

TABLE X
COMPARISON BETWEEN Q-LEARNING WITH ELIGIBILITY TRACES AND OTHER METHODS FOR ACTIVE LOAD DEMAND (PD= 2630 MW)

Method	Cost, $/h	Method	Cost, $/h
PSO [34]	32858	SCA [35]	32867.025
GA [34]	33063.54	APSO [36]	32742.77
SPSO [11]	32798.69	CPSO [34]	32834
PC_PSO [11]	32775.36	BF [34]	32784.5
SOH-PSO[34][11]	32751.39	MDE [34]	32704.9
MTS [33]	32796.13	TSA [33]	32917.87
SA [35]	32786.4	DSPSO-TSA [33]	32715.06
SCA [35]	32867.025	Q-Learning	32676

TABLE XI
OPTIMAL GENERATED POWER AND LOSSES BY Q-LEARNING WITH ELIGIBILITY TRACES (PD=2630 MW)

Unit	P (MW)	Unit	P (MW)	Unit	P (MW)
1	403.7229	6	427.4191	11	45.2892
2	426.7635	7	459.4340	12	55
3	124.6209	8	60	13	25
4	121.7764	9	25	14	15
5	434.7174	10	25	15	16.6673
Total Genration (MW)		2665.4	Loses (MW)	35.4102	
Total cost ($/h)		32676			

VI. CONCLUSION

Q-Learning is used to solve the Economic Dispatch problem with non-convex cost function. Eligibility traces are used to speed up the learning process. The study cases included Economic Dispatch problems with valve point loading effect, multiple fuel options, and transmission losses. Simulation results showed that Q-Learning with eligibility traces achieved the lowest fuel generation cost compared to some Heuristic optimization techniques. The importance of the developed algorithm is that once the learning process is complete, the optimal generated power distribution for any active load demand can be retrieved without any addition efforts unlike other optimization techniques.

APPENDIX

The B loss coefficients are given as follows

[0.0014	0.0012	0.0007	-0.0001	-0.0003	-0.0001	-0.0001	-0.0001	-0.0003	-0.0005	-0.0003	-0.0002	0.0004	0.0003	-0.0001
0.0012	0.0015	0.0013	0.0000	-0.0005	-0.0002	0.0000	0.0001	-0.0002	-0.0004	-0.0004	0.0000	0.0004	0.0010	-0.0002
0.0007	0.0013	0.0076	-0.0001	-0.0013	0.0009	-0.0001	0.0000	-0.0008	-0.0012	-0.0017	0.0000	-0.0026	0.0111	-0.0028
-0.0001	0.0000	-0.0001	0.0034	-0.0007	-0.0004	0.0011	0.0050	0.0029	0.0032	-0.0011	0.0000	0.0001	0.0001	-0.0026
-0.0003	-0.0005	-0.0013	-0.0007	0.0090	0.0014	-0.0003	-0.0012	-0.0010	-0.0013	0.0007	-0.0002	-0.0002	-0.0024	-0.0003
-0.0001	-0.0002	-0.0009	-0.0004	0.0014	0.0016	0.0000	-0.0006	-0.0005	-0.0008	0.0011	-0.0001	-0.0002	-0.0017	0.0003
-0.0001	0.0000	-0.0001	0.0011	-0.0003	0.0000	0.0015	0.0017	0.0015	0.0009	-0.0005	0.0007	0.0000	-0.0002	-0.0008
-0.0001	0.0001	0.0000	0.0050	-0.0012	-0.0006	0.0017	0.0168	0.0082	0.0079	-0.0023	-0.0036	0.0001	0.0005	-0.0078
-0.0003	-0.0002	-0.0008	0.0029	-0.0010	-0.0005	0.0015	0.0082	0.0129	0.0116	-0.0021	-0.0025	0.0007	-0.0012	-0.0072
-0.0005	-0.0004	-0.0012	0.0032	-0.0013	-0.0008	0.0009	0.0079	0.0116	0.0200	-0.0027	-0.0034	0.0009	-0.0011	-0.0088
-0.0003	-0.0004	-0.0017	-0.0011	0.0007	0.0011	-0.0005	-0.0023	-0.0021	-0.0027	0.0140	0.0001	0.0004	-0.0038	0.0168
-0.0002	0.0000	0.0000	0.0000	-0.0002	-0.0001	0.0007	-0.0036	-0.0025	-0.0034	0.0001	0.0054	-0.0001	-0.0004	0.0028
0.0004	0.0004	-0.0026	0.0001	-0.0002	-0.0002	0.0000	0.0001	0.0007	0.0009	0.0004	-0.0001	0.0103	-0.0101	0.0028
0.0003	0.0010	0.0111	0.0001	-0.0024	-0.0017	-0.0002	0.0005	-0.0012	-0.0011	-0.0038	-0.0004	-0.0101	0.0578	-0.0094
-0.0001	-0.0002	-0.0028	-0.0026	-0.0003	0.0003	-0.0008	-0.0078	-0.0072	-0.0088	0.0168	0.0028	0.0028	-0.0094	0.1283]

Bo=[-0.0001 -0.0002 0.0028 -0.0001 0.0001 -0.0003 -0.0002 -0.0002 0.0006 0.0039 -0.0017 0 -0.0032 0.0067 -0.0064], Boo=0.0055.

ACKNOWLEDGMENT

This work was supported by NSF grant ECCS-1128050, ARO grant W91NF-05-1-0314, AFOSR grant FA 9550-09-1-0278, China NNSF grant 61120106011, and China Education Ministry Project 111 (No.B08015)

REFERENCES

[1] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*. New York: Wiley, 1996.
[2] S. Duman, U. Guvenc, and N. Yorukeren, "Gravitational search algorithm for economic dispatch with valvepoint effects," *Int. Rev. Elect. Eng.*, vol. 5(6), pp.2890–2895, 2010.

[3] N. Amjady, and H. Nasiri-Rad, "Economic dispatch using an efficient real-coded genetic algorithm," *IET Gen. Trans. Dist.*, vol. 3(3), pp. 266-278, 2009.
[4] N. Amjady, and H. Nasiri-Rad, "Solution of nonconvex and non smooth economic dispatch by a new Adaptive Real Coded Genetic Algorithm," *Exp. Sys. with Appl.*, pp.5237-5239e45, 2010.
[5] D. Lukman, K. Walshe, and T. R. Blackburn, "Loss Minimisation in Industrial Power System Operation," *Australasian Universities Power Engineering Conference (AUPEC2000)*, Brisbane, pp. 15–20, 2000.
[6] W. M. Lin, and S. J. Chen, "Bid-based dynamic economic dispatch with an efficient interior point algorithm," *Elec. Pwr. and Enr. Sys.*, vol. (24), pp.51-57, 2002.
[7] J. Nanda, D. P. Kothari, and K. S. Lingamurthy, "Economic-emission load dispatch through goal programming techniques," *IEEE Trans. on Enr. Conv.,* vol. 3(1), pp. 26-32, 1998.
[8] G. P. Granelli, and M. Montagna, "Security-constrained economic dispatch using dual quadratic programming," *Elect. Pwr. Syst. Res.,* vol. (56), pp. 71-80, 2000.
[9] H. T. Yang, P. C. Yang, and C. L. Huang, "Evolutionary programming based economic dispatch for units with non-smooth fuel cost functions," *IEEE Trans. on Pwr. Sys.*, vol. 11(1), pp. 112-118, 1996.
[10] L. S. Coelho, and V. C. Mariani, "Combining of chaotic differential evolution and quadratic programming for economic dispatch optimization with valve-point effect," *IEEE Trans. Pwr. Sys.,* vol. 21(2), pp. 989–96, 2006.
[11] K. T. Chaturvedi, M. Pandit, and L. Srivastava, "Self-organizing hierarchical particle swarm optimization for nonconvex economic dispatch," *IEEE Trans. on Pwr. Sys.*, vol. 23(3), pp. 1079-87, 2008.
[12] K. P. Wong, and C. C. Fung, "Simulated Annealing based Economic Dispatch algorithm," *IEE Proc., C 140*, vol. 6, pp. 509 – 515, 1993.
[13] W. M. Lin, F. S. Cheng, and M. T. Tsay, "An improved Tabu search for economic dispatch with multiple minima," *IEEE Trans. on Pwr. Sys.*, vol. 17(1), pp. 108-112, 2002.
[14] M. Vanitha and K. Thanushkodi, "An Efficient Technique for Solving the Economic Dispatch Problem using Biogeography Algorithm," *European J of Scientific Res*., vol. 50(2), pp. 165-172.
[15] S. Sen and G. Weis, "Learning in multi-agent systems, in Multi-agent Systems: A Modern Approach to Distributed Artificial Intelligence," Ed. Cambridge, MA: MIT Press, pp. 259–298, 1999.
[16] R. S. Sutton and A. G. Barto, *Reinforcement Learning–An Introduction*. Massachusetts: Cambridge, MIT Press, 1998.
[17] P. J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavior Sciences*. PhD Thesis, 1974.
[18] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling. Handbook of Intelligent Control," Ed. D.A. White and D.A. Sofge, New York: Van Nostrand Reinhold, 1992.
[19] I. Ahammed, E. A. Jasmin, F. R. Pazheri, and E. A. Al-Ammar, "Reinforcement Learning Solution to Economic Dispatch Using Pursuit Algorithm," *6th IEEE-GCC Conference and Exhibition,* Dubai, UAE, pp. 19-22, 2011.
[20] E. A. Jasmin, I. Ahamed, and V. P. Jagathyraj, "A Reinforcement Learning algorithm to Economic Dispatch considering transmission losses," *Proceedings of TENCON*, 2008.
[21] I. Ciornei, and E. Kyriakides, "A GA – API solution for the economic dispatch of generation in power system operation," *IEEE Trans. on Pwr. Sys.*, pp. 1-9, 2011.
[22] C. L. Chinag, "Improved genetic algorithm for power economic dispatch of units with valve-point effects and multiple fuels," *IEEE Trans. Pwr. Sys.,* vol. 20(4), pp. 1690–1699, 2005.
[23] L. Busoniu, R. Babuska, and B. De Schutter, "Multi-agent reinforcement learning: A survey," *Proceedings of the 9th International Conference on Control, Automation, Robotics and Vision*, Singapore; pp. 527-532, 2006.
[24] Z. L. Gaing, "Particle Swarm Optimization to Solving the Economic Dispatch Considering the Generator Constraints," *IEEE trans. on pwr. sys*., vol. 18(3), pp. 1187–1195, 2003.
[25] U. Guvenc, S. Duman, B. Saracoglu, and A. Özturk, "A Hybrid GA-PSO Approach Based on Similarity for Various Types of Economic Dispatch Problems," *Kaunas University of Technology, Electronics And Electrical Engineering*, vol. 2(108), pp. 109-114, 2011.
[26] A. Bhattacharya and P. K. Chattopadhyay, "Solving complex economic load dispatch problems using biogeography-based optimization," *Expert Sys. with Appl.*, vol. 37, pp. 3605-3615, 2010.
[27] S. R. Rayapudi, "An Intelligent Water Drop Algorithm for Solving Economic Load Dispatch Problem," *Int. J of Elect. and Elect. Eng,* vol. 5(2), pp. 43-49, 2011.
[28] C. E. Lin and G. L. Viviani, "Hierarchical economic dispatch for piecewise quadratic cost functions," *IEEE Trans. Pwr. Appar. Syst*., vol. 103(6), pp. 1170–1175, 1984.
[29] J. H. Park, Y. S. Kim, I. K. Eom, and K. Y. Lee, "Economic load dispatch for piecewise quadratic cost function using Hopfield neural network," *IEEE Trans. Pwr. Sys.*, vol. (8), pp. 1030–1038, 1993.
[30] K. Y. Lee, A. Sode-Yome, and J. H. Park, "Adaptive Hopfield neural network for economic load dispatch," *IEEE Trans. Pwr. Sys.,* vol. 13, pp. 519–526, 1998.
[31] T. Jayabarathi and G. Sadasivam, "Evolutionary programming based economic dispatch for units with multiple fuel options," *Eur. Trans. Electr. Pwr.*, vol. 10(3), pp. 167–170, 2000.
[32] P. S. Manoharan, P. S. Kannan, S. Baskar, and M. W. Ruthayarajan, "Penalty parameter-less constraint handling scheme based evolutionary algorithm solutions to economic dispatch," *IET Gener. Trans. Distrib.,* vol. 2(4), pp. 478–490, 2008.
[33] S. Khamsawang and S. Jiriwibhakorn, "DSPSO-TSA for economic dispatch problem with nonsmooth and noncontinuous cost functions," *Energy Conversion and Management*, vol. 51(2), pp. 365-75, 2010.
[34] J. B. Park, Y. W. Jeong, J. R. Shin, K. Y. Lee KY, "An improved particle swarm optimization for nonconvex economic dispatch problems," *IEEE Transactions on Power Systems,* vol. 25(1), pp. 156-166, 2010.
[35] A. Selvakumar and T. Khanushkodi, "Optimization using civilized swarm: solution to economic dispatch with multiple minima," *Elect. Pwr. Sys. Res.,* vol. 79(1), pp. 8-16, 2009.
[36] B. K. Panigrahi and V. R. Pandi, S. Das, "Adaptive particle swarm optimization approach for static and dynamic economic load dispatch," *Ener. Convr. and Mang.*, vol. 49(6), pp. 1407-15, 2008.