Correlated Equilibrium Q-learning for Multi-objective Reactive Power Optimization Considering Grid Side Carbon Emissions

Hong Hu, Wenmei Wu
Power Distribution Dept, Xingyi Power Supply Bureau
Xingyi, China
951445945@qq.com

Min Tan, Shaohua Xiao, Chuanjia Han
School of Electric Power Engineering, South China
University of Technology, Guangzhou, China
tanminscut@126.com

***Abstract*—In order to meet the development trend of smart grid, the correlated equilibrium Q-learning (CEQ) algorithm is proposed for multi-regional reactive power optimization. Meanwhile, in response to the national strategy of low carbon environmental protection, CO_2 emission is considered as one of the control objectives in reactive power optimization. In this paper, CEQ algorithm is adopted to allocate the control variables rationally, through the correlated equilibrium game among areas and information communication and sharing to achieve multi-regional reactive power optimization, which solves the limited information-sharing mechanisms and curse of dimensionality problem effectively. Simulation of the IEEE 9-bus system indicates that through the combine of pre-learning and online learning CEQ algorithm solves the multi-regional collaborative reactive power optimization quickly and rationally.**

Keywords-multi-regional reactive power optimization; low-carbon electricity; correlated equilibrium; reinforcement learning

I. INTRODUCTION

Reactive power optimization is to control the system voltage and reactive power distributions to achieve a better system performance. Mathematically, reactive power optimization is a nonlinear mixed integer programming problem [1].

With a lot of intermittent renewable energy sources connected to the power grid, the power system is becoming a complex non-linear large-scale system, which brings about the problems of coordination interaction, massive data and communication bottlenecks [2]. Because of the difficult in information exchange and high cost caused by wide geographical distribution of the power grid, not all the information can be sent to the centralized control center in a centralized decision-making way. While, the existing reactive power optimization algorithms are not suitable to deal with the problems of curse of dimensionality, coordination and interaction ,which include classical optimization algorithm and part of modern optimization algorithm such as genetic algorithms [3], quantum genetic algorithm [4], PSO algorithm [5] and ant immune algorithm [6]. Meanwhile the global grid information is needed to run centralized optimization using these algorithms. Therefore, the reactive power optimization algorithm that is applicable to the new situation need to be proposed.

In addition, with global warming attracting more and more attention, "low-carbon life" concept are gradually accepted. Since a large part of China's carbon dioxide (CO_2) emission comes from the consumption of fossil fuels in power industry. Thus, research on low-carbon electricity techniques is meaningful, such as low-carbon power system planning and operation [7], low-carbon electricity dispatch [8], the power plant carbon capture and storage technologies [9]. While these research is mainly from the perspective of generation side, so this paper introduces grid side carbon emission into the objective function of reactive power optimization model.

Since not all the information can be sent to the control center, the collection of reactive power optimization information is restricted, which can be solved through regional "autonomy" and interregional "coordination". Therefor correlated Q-learning (CEQ) algorithm for multi-regional collaborative reactive power optimization is proposed. The algorithm realized through information exchange of state-action value function matrix and cooperative game among regions.

II. GRID SIDE CO_2 EMISSION MODEL

The grid-side CO_2 emission flow is similar to power flow. The difference between the two flows is that the existence of grid-side carbon emission flow is virtual network flow which depends on power flow and can be understood as carbon labeling of branch power flow.

Consider power grid N with n buses, s generators buses, b branches and L_{ij} denotes the branch connecting bus i and bus j .

For lossless network $P_{ij=}P_{ji}$, then the active power flow from generator bus k to bus i can be described as:

$$P_{ik} = \alpha_{ik} \cdot P_{sk} \tag{1}$$

Where P_{sk} is the output active power of generation bus k, α_{ik} represents the contribution rate of active power from generator bus k over bus i. Specific derivation can be found in[10].

Similarly, from the point of the composition of the total active power flowing into bus $i,$ the proportion of the active power from generator bus k to bus i in total active power flowing into bus i can be represented as follows:

$$\beta_{ik} = \frac{P_{ik}}{P_i} = \frac{\alpha_{ik} \cdot P_{sk}}{P_i} \qquad (2)$$

Where P_i is the total active power of bus *i*.

For lossy network $P_{ij\neq} P_{ji}$, the power loss ΔP_{ij} of branch L_{ij} can be expressed as:

$$\Delta P_{\text{ij}} = g_{ij}[V_i^2 + V_j^2 - 2V_iV_j \cos\theta_{ij}] \qquad (3)$$

Where V_i and V_j respectively denote voltage amplitudes of bus i and bus j; θ_{ij} is the voltage angle difference between buses i and j, g_{ij} represents the conductance of branch L_{ij} .

According to the principle of proportional-sharing[11], β_{ij} is also the power loss contribution rate of bus *k* to branch L_{ij}. Hence the power loss of branch L_{ij} can be described as:

$$\Delta P_{\text{ij}} = \sum_{k=1}^{S}(\beta_{ij} \cdot \Delta P_{\text{ij}}) = \sum_{k=1}^{S}((\frac{\alpha_{ik} \cdot P_{sk}}{P_i}) \cdot \Delta P_{\text{ij}}) \qquad (4)$$

Where *s* is the number of generator buses.

The carbon emission intensity is numerically equal to the generator carbon emission factor multiplying the active power output. The analogy can be obtained that the ΔC_{ij}, which is the carbon emission loss intensity of branch $L_{ij,}$ is formulated as:

$$\Delta C_{\text{ij}} = \sum_{k=1}^{S}(\beta_{ik} \cdot \Delta P_{\text{ij}} \cdot \delta_{ik}) = \sum_{k=1}^{S}((\frac{\alpha_{ik} \cdot P_{sk}}{P_i}) \cdot \Delta P_{\text{ij}} \cdot \delta_{ik}) \qquad (5)$$

Where δ_{ik} denotes CO_2 emission factor of generator bus *k*.

III. Mathematical Model Of Multi-Objective Reactive Power Optimization

Multi-objective reactive power optimization model includes objective function and constraints which mainly include controlled variable constraints, state variable constraints and power flow constraints.

A. *Objective Function*

The multiple objectives include the grid side CO_2 emission loss, power loss and voltage stability component.

From an environmental point of view, the minimum grid side CO_2 emission loss can be rewritten as:

$$\min(\sum_{i,j\in N_L}(\Delta C_{\text{ij}})) = \sum_{i,j\in N_L}(\sum_{k=1}^{S}((\frac{\alpha_{ik} \cdot P_{sk}}{P_i}) \cdot \Delta P_{\text{ij}} \cdot \delta_{ik})) \qquad (6)$$

Where ΔC_{ij} is the CO_2 carbon emission loss of branch L_{ij}, P_{sk} is the output active power of generator bus *k*; P_i is the total active power flowing into bus *I*,ΔP_{ij} is the power loss of branch L_{ij} , $\boldsymbol{N_L}$ is the set of branches.

From the economy point of view, the minimum power loss can be expressed as follows:

$$\min(\sum_{i,j\in N_L}(\Delta P_{\text{ij}})) = \sum_{i,j\in N_L}(g_{ij}[V_i^2 + V_j^2 - 2V_iV_j \cos\theta_{ij}]) \qquad (7)$$

Where ΔP_{ij} is the active power loss of branch L_{ij}.

From the safety point of view,to make the synthesized voltage stability of buses in the system optimal, the objective function is:

$$\min(\sum_{i=1}^{n}|\Delta V_i|) = \sum_{i=1}^{n}\left|\frac{2V_i - V_{i\max} - V_{i\min}}{V_{i\max} - V_{i\min}}\right| \qquad (8)$$

Where ΔV_i is the voltage stability component of bus *i*, V_{imax} and V_{imin} are the maximum voltage limits and minimum voltage limits of bus *I*, *n* is the number of the buses.

Considering the grid side CO2 emission loss, branch loss and voltage stability component, the multi-objective function can be represented as follows:

$$F = \min(\lambda_1 \cdot \sum_{i,j\in N_L}(\Delta C_{\text{ij}}) + \lambda_2 \cdot \sum_{i,j\in N_L}(\Delta P_{\text{ij}}) + \lambda_3 \cdot \sum_{i=1}^{n}|\Delta V_i|) \qquad (9)$$

Where λ_1 is the grid side CO_2 emission loss weighting factor; λ_2 is the power loss weighting factor; λ_3 is the voltage stability component weighting factor. The values of these weighting factors meet $\lambda_1\in(0,1]$,$\lambda_2\in(0,1]$,$\lambda_3\in(0,1]$ and $\lambda_1+\lambda_2+\lambda_3=1$.

B. *Constraints*

Reactive power optimization constraints involve controlled variable constrained, state variable constraints and power flow constraints. We choose the capacity of shunt capacitance Q_c and transformer ratio K_T as the controlled variable constrains. The controlled variables should meet following conditions:

$$\begin{aligned} &Q_{Ci}^{\min} \le Q_{Ci} \le Q_{Ci}^{\max}, i \in N_{\text{C}} \\ &K_{Tj}^{\min} \le K_{Tj} \le K_{Tj}^{\max}, j \in N_K \end{aligned} \qquad (10)$$

Where N_c represents the set of buses with adjustable reactive power capacity; N_k is the set of branches with adjustable transformer ratio.

The state variables include the output active power P_G and reactive power Q_G of the generator and the bus voltage *V*. The state variables should meet following conditions:

$$\begin{aligned} &P_{\text{G}i}^{\min} \le P_{\text{G}i} \le P_{\text{G}i}^{\max}, i \in N_{\text{G}} \\ &Q_{\text{G}j}^{\min} \le Q_{\text{G}j} \le Q_{\text{G}j}^{\max}, j \in N_{\text{G}} \\ &V_k^{\min} \le V_k \le V_k^{\max}, k \in N_B \end{aligned} \qquad (11)$$

Where $\boldsymbol{N_G}$、$\boldsymbol{N_B}$ represent the set of generator buses and the set of buses.

Active power balance constraints and reactive power balance constraints are included in the power flow constraints equation. It can be expressed as follows:

$$\begin{aligned} & P_{\mathrm{G}i} - P_{\mathrm{D}i} - V_i \sum_{j \in N_i} V_j (g_{ij} \cos\theta_{ij} + b_{ij} \sin\theta_{ij}) = 0 \\ & Q_{\mathrm{G}i} - Q_{\mathrm{D}i} - V_i \sum_{j \in N_i} V_j (g_{ij} \sin\theta_{ij} + b_{ij} \cos\theta_{ij}) = 0 \end{aligned} \tag{12}$$

Where P_G、Q_G denote the output active and reactive power of the generator bus, P_D、Q_D represent the active and reactive loads of the load buses; g_{ij}、b_{ij} are the conductance and the susceptance of the branch L_{ij}.

IV. Correlated Q-Learning For Multi-Regional Reactive Power Optimization

Greenwald first proposed CEQ algorithm [12]. Correlated equilibrium is a probability distribution of the joint action space. The correlated equilibrium policy is that each agent selects its action according to the conditional probability of other agents over it to maximum its accumulation of reward value[13].

The eligibility trace update policy of agent i is as follows:

$$e_{i,k}(s,\vec{a}) = \begin{cases} \gamma\lambda e_{i,(k-1)}(s,\vec{a}) + 1 & (s,\vec{a}) = (s_k,\vec{a}_k) \\ \gamma\lambda e_{i,(k-1)}(s,\vec{a}) & \text{otherwise} \end{cases} \tag{13}$$

Where $e_{i,k}$(s,**α**) is the eligibility trace of agent i in the kth iteration for the joint action **α**;γ is the discount factor, 0≤γ≤1; λ is the attenuation factor, 0≤λ≤1; s represents state.

The agent i 's Q matrix update can be formulated as follows:

$$\begin{cases} Q_{i,(k+1)}(s,\vec{a}) = Q_{i,k}(s,\vec{a}) + \alpha\delta_{i,k} e_{i,k}(s,\vec{a}) \\ Q_{i,(k+1)}(s_k,\vec{a}_k) = Q_{i,(k+1)}(s_k,\vec{a}_k) + \alpha\rho_{i,k} \\ \delta_{i,k} = R_i(s_k,s_{k-1},\vec{a}_k) + \gamma Q_{i,k}(s_{k+1},\vec{a}_g) - Q_{i,k}(s_k,\vec{a}_g) \\ \rho_{i,k} = R_i(s_k,s_{k-1},\vec{a}_k) + \gamma Q_{i,k}(s_{k+1},\vec{a}_g) - Q_{i,k}(s_k,\vec{a}_k) \end{cases} \tag{14}$$

Where R_i is the agent i 's immediately reward function, α is the learning factor,0≤α≤1, $\boldsymbol{\alpha}_{\mathbf{g}}$ denotes the joint greedy action policy.

The reward function can be described as follows:

$$R = -(\lambda_1 \cdot \sum_{i,j \in N_L} (\Delta C_{\mathrm{ij}}) + \lambda_2 \cdot \sum_{i,j \in N_L} (\Delta P_{\mathrm{ij}}) + \lambda_3 \cdot \sum_{i=1}^{n} |\Delta V_i|) \tag{15}$$

Where ΔC is theCO$_2$ emission loss, ΔP is the branch loss and ΔV is the voltage stability component.

Through solving linear programming the optimal action probability distribution can be get. The objective and constraints of linear programming is represented as follows:

$$\begin{cases} f = \max \sum_{i \in N} \sum_{a \in A(s)} \pi_s(\vec{a}) Q_i(s,\vec{a}) \\ \sum_{a_{-i} \in A_{-i}} \pi_s(a) Q_i\left(s,(a)\right) \geq \sum_{a_{-i} \in A_{-i}} \pi_s(a) Q_i\left(s,(a_{-i},a'_i)\right) \\ \sum_{a \in A(s)} \pi_s(a) = 1, 0 \leq \pi_s(a) \leq 1, i \in N, s \in S \end{cases} \tag{16}$$

Where index $-i$ represents the set of other agents except agent I; α_i' denotes any other action of agent i apart from α_I; π is the action probability. ***N***、***S***、***A(s)*** are the set of agents, the set of states and the set of action.

Power system reactive power optimization generally adopts the policy of balance on the spot, which is obvious regional. Hence the power grid can be divided into several regions. Every controlled variable in reactive power optimization algorithm respect an agent in CEQ algorithm. The Q values of each agent can be observed by any other agents. Meanwhile the agent can solve the correlated equilibrium independently and select the action.

The framework of regional reactive power optimization describes in Figure 1

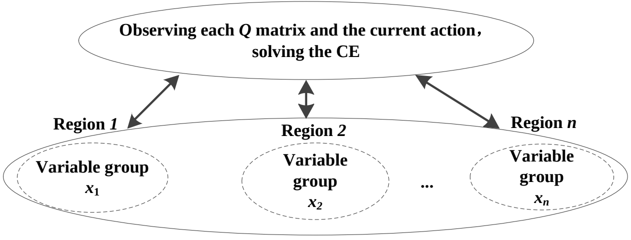

Figure 1. Framework chart of regional reactive power optimization

At the beginning of reactive power optimization, before the power flow calculation the values of controlled variables are determined. Next, run the power flow calculation and calculate the reward function and state value with the power flow result. Then, using CEQ algorithm to solve reactive power optimization and calculate the value of controlled variables. Finally loop until the best action value is acquired. The flow chart of reactive power optimization is shown in Figure 2

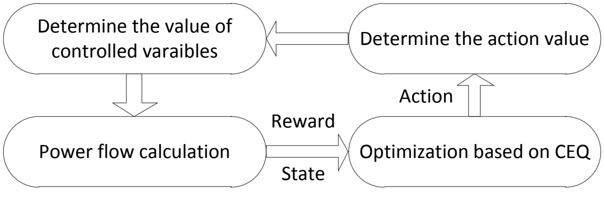
Determine the value of controlled varaibles
Determine the action value
Action
Reward
Power flow calculation
Optimization based on CEQ
State

Figure 2. Reactive power optimization flowchart

To sum up, Figure 3 shows the flow chart of multi-regional reactive power optimization algorithm based on CEQ.

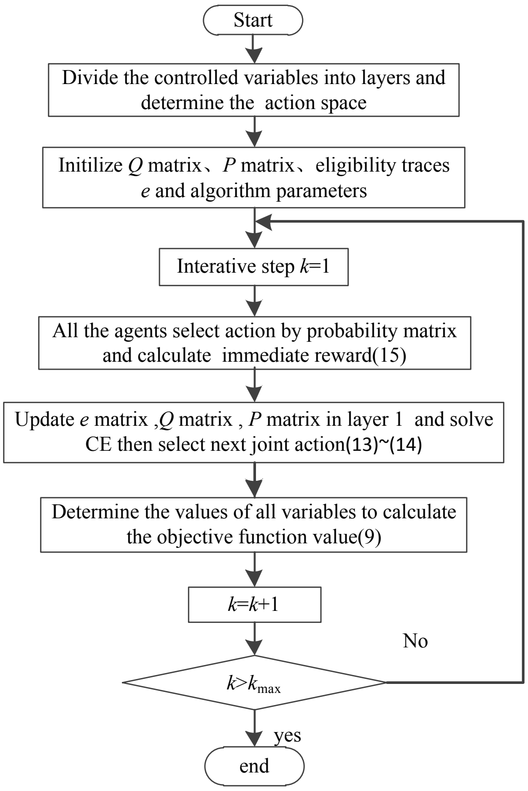
Start
Divide the controlled variables into layers and determine the action space
Initilize Q matrix、P matrix、eligibility traces e and algorithm parameters
Interative step k=1
All the agents select action by probability matrix and calculate immediate reward(15)
Update e matrix ,Q matrix , P matrix in layer 1 and solve CE then select next joint action(13)~(14)
Determine the values of all variables to calculate the objective function value(9)
k=k+1
No
k>k_{max}
yes
end

Figure 3. Algorithm flowchart

V. EXAMPLES AND SIMULATION ANALYSIS

A. Simulation Model

In this paper, the simulation test is carried out by IEEE9 Bus power system on Matlab7.10 simulation platform and the power flow calculation is based on Matpower4.1. There are 3 generators,3 load buses and 9 branches in the IEEE9 bus system. The system is divided into 3 regions according to the distribution and connection of the buses. The bus partition is shown in figure4:

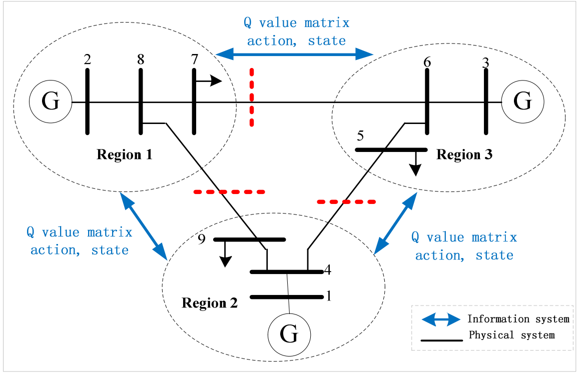

Figure 4. Bus partition in IEEE9 system

As is shown on the picture above, the grid is partitioned in accordance with the red dotted line. The various regions communicate and share information with each other in the reactive power optimization process based on CEQ. The capacities of reactive power compensation in buses 5,7,9 are chosen as controlled variables.

B. Simulation Result s

Q-learning is process of trial and error. In the early stages, Q-learning is relatively random and blind. It's not appropriate to be directly applied to online optimization of the actual system. So we need to get pre-learning Q value matrix with learning experience, and then carry on online learning.

After a lot of simulation analysis, for pre-learning process ,the value of learning factor, discount factor and eligibility factor are respectively 0.94,0.10,0.50; for online learning process, the value of learning factor, discount factor and eligibility factor are respectively 0.40, 0.10,0.30.

In response to the policy of low-carbon electricity, the grid side carbon emissions loss is considered as one of reactive power optimization objectives. The carbon emission factor of generators in IEEE9 buses system is shown in Table I:

TABLE I. CARBON EMISSION INTENSITY OF IEEE 9 BUS

Bus	Generator type	δ (kg/kW·h)
Bus 1	Thermal power (coal-fired)	1.01
Bus 2	Thermal power (coal-fired)	0.95
Bus 3	Hydropower	0

Figure 5 (a) shows the convergence curve of the objective function by pre-learning process. figure 5 (b) shows the convergence curve of the objective function by online learning.

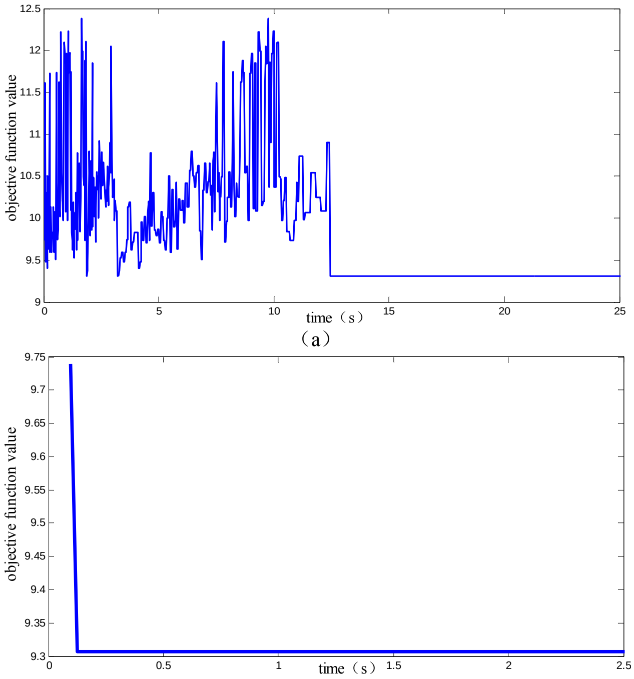

Figure 5. Convergence curve of objective function

As is shown in Figure5 (a), the objective function value of CEQ algorithm can converge to a minimum after a series of trial and error process. But it takes a lot of time, which cannot meet the power demand in time. From Figure 5 (b) after the pre-learning the convergence process of the algorithm becomes faster and more stable. The convergence time is reduced from12.43s to 0.13s. After the CEQ optimization simulation, the reactive power configuration in IEEE9 buses system is shown in TABLE Ⅱ.

TABLE II. REACTIVE POWER COMPENSATION CAPACITY

Compensation bus	Bus 5	Bus 7	Bus 9
Input capacity Q(Mvar)	24	28	40

The branch loss, voltage stability component and carbon emission lossof IEEE9 system before and after the reactive power optimization are compared in TABLE III.

TABLE III. RESULTS COMPARISON BEFORE AND AFTER REACTIVE POWER OPTIMIZATION

	Branch loss（MW）	Voltage Stability Component	CO_2 emission loss（kg）	Objective function
Before optimization	4.95	2.03	3.27	10.25
After optimization	4.81	1.34	3.15	9.3

From the chart above, it is obvious that all the indicators have been improved. Among them, the voltage quality indicators improved the most, with the performance indicator increasing by 33.99%.More over branch loss value decreased by 2.83%, carbon emissions decreased by 3.67% and total objective function decreased by 9.26 %.

To validate the algorithm, the results obtained by 100 times simulations of Q-learning algorithm and CEQ algorithm are shown in TABLE Ⅳ.

TABLE IV. ANALYSIS OF SIMULATION RESULTS

Algorithm	Average convergence time（s）	The minimum of objective function	The varianceof objective function	Capacity of compensation (MVar)
Q	0.68	9.307	0.0000	92
CEQ	0.22	9.307	0.0103	92

As table.4 shows, CEQ algorithm obtains optimal solution faster than Q-learning algorithm. The average convergence time of CEQ algorithm is 0.22s while the average convergence time of Q-learning algorithm is 0.68s. But the convergence stability of CEQ algorithm is a bit worse. So the CEQ algorithm has a short convergence time and good convergence stability. It can be applied to online multi-section extensions.

After the adoption of the pre-learning, the CEQ algorithm can achieve fast optimization. To further test the performance of the algorithm, we change the load section to carry on the test. figure 6 shows the convergence curves of objective function of Q and CEQ when the system load increases by 8%.

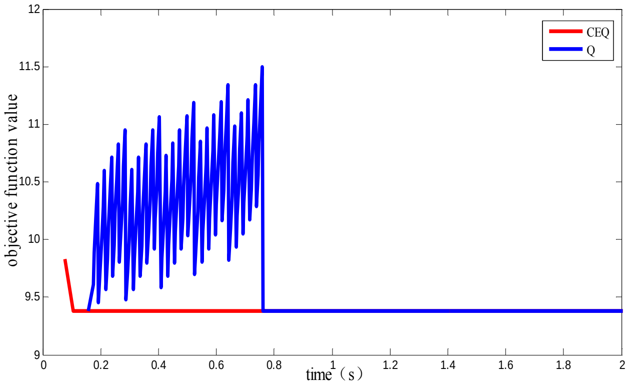

Figure 6. Convergence curve of each algorithm

With the load fluctuating slightly, CEQ algorithm still can converge quickly to the minimum by the correlated equilibrium game between regions. Therefore, the algorithm can be further applied to the dynamic load reactive power optimization problem with a promising application.

VI. CONCLUTION

Multi-regional reactive power optimization model based on CEQ cooperative algorithm is proposed in this paper, meanwhile CO_2 emission loss is considered as one of the control objectives which are conducive to low-carbon environment. Without knowing the global information of power grid, CEQ cooperative algorithm can also solve the communication bottleneck problem through correlated equilibrium game among areas. Simulation shows that that the algorithm can effectively find the optimal solution, and has a faster convergence speed. Thus, the algorithm can deal with the development trend of smart grid which has good prospects.

REFERENCES

[1] K. R. C. Mamandur, R. D. Chenoweth, “Optimal control of reactive power flow for improvements in voltage profiles and for real power loss minimization” [J] Power Engineering Review IEEE, vol. PAS-100, no. 7, pp. 29-30, July. 1981.

[2] C. Zhao, W. Wu, B. Zhang, “Multi-agent based software platform for new generation of EMS” [J] Automation of Electric Power Systems, vol. 33, no. 13, pp. 47-52, 2009.

[3] Suresh R, Kumarappan N. “Genetic algorithm based reactive power optimization under deregulation”[C] ICTES. IET-UK International Conference on. IET, pp. 150-155, 2007.

[4] Saitoh A, Rahimi R, Nakahara M. “A quantum genetic algorithm with quantum crossover and mutation operations” [J]. Quantum Information Processing, vol. 13, no. 3 , pp. 737-755, 2013.

[5] Zhong-Yan L U, Deng J X, Wang Y H. “Reactive Power Optimization Based on Particle Swarm Optimization Algorithm With Immunity”[J]. Power System Technology, vol. 32, no. 24 , pp. 55-59, 2008.

[6] Watkins A, Timmis J, Boggess L. “Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Learning Algorithm”[J]. Genetic Programming & Evolvable Machines, vol. 5, no. 3, pp. 291-317, 2004.

[7] Q.X.Chen,C.Q.Kang,Q.Xia.”Mechanism and Modell ing Approach to Low-carbon Power Dispatch”[J], Automation of Elect ric Power Systems, vol. 34, no. 12, pp. 18-23, 2010.

[8] Kang Chongqing, Chen Qixin, Xia Qing. “Prospects of low-carbon electricity”[J]. Power System Technology, vol. 3, no. 2, pp. 1-7, 2009.

[9] Xu Ruina，Chen Wenying，Wu Zongxin．Cost and performance of power plants with CO2 capture [J]．Journal of Tsinghua University ：Science and Technology，vol.49, no. 9, pp. 103-106,2009.

[10] Li B, Song Y, Hu Z. “Carbon Flow Tracing Method for Assessment of Demand Side Carbon Emissions Obligation ”[J]. Sustainable Energy IEEE Transactions on, vol. 4 , no. 4, pp. 1100-1107, 2013.

[11] Zhou Tianru，Kang Chongqing，Xu Qianyao，et al．Preliminary theoretical investigation on power system carbon emission flow[J]． Automation of Electric Power Systems， vol. 36 , no. 7, pp. 38-43 2012.

[12] A. R. Greenwald, K. Hall, “Correlated Q learning,” Proceedings of the Twentieth International Conference on Machine Learning, pp. 242-249,2002.

[13] T. Yu, B. Zhou, K. W. Chan, “Stochastic optimal relaxed automatic generation control in non-markov environment based on multi-step Q(λ) learning,” IEEE Transactions on Power Systems, vol. 26, no. 3, pp. 1272-1282, 2011.