Risk-based Reactive Power Optimization Based on Tribe Q-Learning Algorithm

Li Feng[1], Xu Zhibin[1], Xiao Li[1]

[1]Guangdong Electric Power Research Institute of energy technology limited liability company. China.
lifeng186@126.com

Abstract. In this paper, the risk assessment theory is introduced into the traditional reactive power optimization problem. Moreover, a novel tribe Q-learning algorithm with knowledge transfer is proposed, which is developed from the search mechanism of artificial intelligence algorithm and the iteration mode of Q-learning. The Q matrix is adopted as the knowledge matrix for the storage of the search information of the tribe. During online learning, the rate of TQL can be accelerated significantly via the knowledge transfer. The simulation on IEEE 118-bus systems demonstrates that the rate of TQL is two to twenty times faster than that of other AI algorithms while the global convergence can be ensured.

1. Introduction

With the development of the industrialization process, the construction of power system have been accelerated. The development of interconnection of regional power grid and cross-regional transmission has become faster. At the same time, more large-scale wind energy, solar energy and electric vehicles has connected to the system in distribution network side, which makes the power grid become more complex and may result in severe challenges to the secure and stable operation of power grids. In order to obtain an appropriate trade-off between system security and economy, since the 1990s, several scholars have studied the security issues of reactive power optimization[1]. Based on the planning and adopting the traditional reactive power optimization model, reference [2] attempted to make up for the deficiencies of traditional methods through the rational configuration of reactive power compensation location. Reference [3] has proposed a reactive power optimization model based on Monte Carlo simulation and voltage security constraint, which takes the improvement of the node voltage level as the goal of optimization. From the perspective of risk, reference [4] has analysed the influence of power loss, voltage instability and voltage violations on the operation of power system, and configured the various reactive power resources in the system with the aim of minimizing operation risk. However, the above studies evaluate the security of the system from the perspective of static voltage state, ignoring the effects of line overload and the effects of load fluctuations.

In order to improve the ability of power system to withstand the operation risks, the theory of power system risk assessment is introduced into the traditional reactive power optimization issue, and a risk-based reactive power optimization (RBRPO) mathematical model is constructed. The model aims at reducing operational risk and active power loss of the system and adopts the probability model to evaluate the risk of transmission line overload and node voltage violation when the transmission line occur faults at the same time to reduce the operational risk and the active power loss of the system.

Risk-based reactive power optimization is a complex mixed discrete nonlinear programming issue. The solution to this kind of issues mainly includes classical mathematical method and heuristic

artificial intelligence algorithm. Compared to classical mathematical method, artificial intelligence (AI) algorithms such as genetic algorithm (GA) [5], particle swarm optimization (PSO) [6], artificial bee colony (ABC) [7] and so on have been widely applied to various areas of power system optimization due to its outstanding features of less dependence of an accurate system model, convenience of application and global optimization and its suitability for dealing with discrete, non-linear large-scale issues [8] [9]. However, these algorithms can only deal with the problem in isolation without the ability to store information and self-learning, which results in inefficiencies in dealing with similar tasks.

Nowadays transfer technology has become a powerful tool to accelerate the process of machine learning for similar multitasking optimization. In practical work, many historical task and new tasks being executed have a number of common features in essence. Transfer learning is to find the similarity between the past and present, using the previous knowledge to guide the current task, which significantly improve the efficiency of task optimization [10]. Based on the previous analysis, this paper introduces a brand new tribe Q-learning algorithm (TQL) with knowledge transfer to solve the risk-based reactive power optimization model. Different from the common artificial intelligence algorithm of random search mode, TQL uses four kinds of individuals, i.e., tribal chiefs, civilians and rangers as the search subject to find the optimal solution. TQL uses the Q matrix in the Q-learning algorithm to store the group optimization information and guide the next step of the optimization method. At the same time, the dimensionality of the knowledge matrix has been reduced, which avoids the curse of dimensionality in the large-scale system. In the pre-learning process, TQL stores the optimization information of the source task in the optimal knowledge matrix. Through the extraction of the similarity, the initial matrix of the similar task is formed in the form of non-linear transfer. Therefore, the optimization process for new tasks will be significantly accelerated during the online learning process. In order to verify the effectiveness of the new algorithm, TQL is applied for RBRPO of 24 scenarios on IEEE118-bus system, which performance is compared with that of other existing AI algorithms.

2. Mathematical model of RBRPO

2.1. Operation risk assessment

The operation risk assessment of power systems means a comprehensive evaluation with the possibility and severity of random perturbations, which can be described by the sum of the product of probability and the consequence of each random disturbance [11]:

$$RI = \sum_i P(s_i) I(s_i) \tag{1}$$

where$s_i$is the ith random perturbations, $P(s_i)$ and $I(s_i)$ are the probability and risk index of s_i, respectively.

This article focused on contingency taking into account the failure of the transmission line outage.According to the statistical data, the failure rate of transmission line L_i at a certain time intervalfollows the Poisson distribution, thus its can be described as

$$P(\mathrm{L}_i) = 1 - \exp(-\lambda_i \Delta t / 8760) \tag{2}$$

whereλ_i is the annualfailure rate of transmission line i;Δt is the fault calculation time, the unit is one hour.

If the line outage is an independent event, the probability of a single failure at any time can be described as[12]:

$$\rho_{\mathrm{L}_i} = P_{\mathrm{L}_i} \prod_{j \in U_{\mathrm{L}}, j \neq i} (1 - P_{\mathrm{L}_j})) \tag{3}$$

where$U_L$is the setof all the normal operational transmission lines, sowe can get the probability of two and more than two lines failure occurred.

The outage of a transmission line may results in the transfer of active power flow, and a sudden line overload or a severe node voltage deviation may happen in the neighbourhood of the failure point.

In order to distinguish the probability of failure more effectively between low probability but serious failure andhigh probability but slight failure, a utility function is employed so as to fully describe the risk index of branch power and node voltage.

The branch power risk index is used to describe the overload of the line power flow, which is defined as follows:

$$RI_t = \sum_{\alpha} a_t (T_i - T_{i\max}) / T_{i\max} \tag{4}$$

where$\alpha$is the set of overload transmission line;$T_i$is the apparent power flowing through line i; $T_{i\max}$ is the limit of line power flow;a,b are positive real numbers.

The node voltage risk indexmeasures the extent of the node voltage overrun, which can be defined as follows:

$$RI_u = \sum_{\beta} a_u (\omega_u + b) \tag{5}$$

$$\omega_u = \begin{cases} \sum_{\beta} \dfrac{U_i - U_{i\max}}{U_{i\max}} & U_i > U_{i\max} \\ \sum_{\beta} \dfrac{U_{i\min} - U_i}{U_{i\min}} & U_i < U_{i\min} \end{cases} \tag{6}$$

where$\beta$is the set of the voltage overrun node; $V_i$is the actual voltage amplitude of node i; $U_{i\max}$ is the upper limit of voltage for node i; $U_{i\min}$ is the lower limit of voltage for node I;a_u,b are positive real numbers.

Taking into account the probability of failure of transmission lines, the system comprehensive risk indexcan be calculated as follows:

$$RI = \sum_{C} \rho_k (RI_t^k + RI_u^k) \tag{7}$$

where$\rho_k$is the kth probability of expected faultoccurrence; $C$is the set of expected fault.

2.2. Objective function andConstraints ofRBRPO

Under the premise to meet the various operational constraints, RBRPO can adjust the distribution of the power flow by reasonably configuring the switching capacity of the reactive power compensation device, the generator terminal voltage and the transformer tap ratio, so as to reduce the active power loss of the power grid and operational risk as much as possible. In this paper, the linear weighting method is adopted to convert multi-objective problem into single objective processing.The objective function of RBRPO can be described as follows:

$$\min f(x) = \omega_1 P_{\text{Loss}}(x) + \omega_2 RI(x) + \omega_3 V_{\text{d}}(x) \tag{8}$$

where$P_{\text{Loss}}(x)$, $RI(x)$ and $V_{\text{d}}(x)$denote the three objectives of the active power loss, risk index and voltage deviation component after normalization, respectively;ω_1, ω_2, $\omega_3$are the weights of each objective.

The voltage deviation componentV_{d} can be calculated as follows:

$$V_{\text{s}} = \sum_{i,j \in S_N} \left| \frac{2U_i - U_{i\max} - U_{j\min}}{U_{i\max} - U_{j\min}} \right| \tag{9}$$

where$S_N$is the set of nodes.

The active power loss P_{Loss}can be described as:

$$V_{\text{s}} = \sum_{i,j \in S_N} G_{ij} [U_i^2 + U_j^2 - 2U_i U_j \cos\theta_{ij}] \tag{10}$$

Where$\theta_{ij}$is the voltage phase angle difference between nodes i and j;$G_{ij}$is the conductance ofline$i$-j.

Constraints include power flow constraints, control variable constraints, and state variable constraints[13]:

$$\begin{cases} P_{Gi}-P_{Di}=U_i\sum_{j\in i}U_j(G_{ij}\cos\theta_{ij}+B_{ij}\sin\theta_{ij}) \\ Q_{Gi}-Q_{Di}=U_i\sum_{j\in i}U_j(G_{ij}\sin\theta_{ij}-B_{ij}\cos\theta_{ij}) \\ Q_{Ci\min}\le Q_{Ci}\le Q_{Ci\max},\ i\in S_C \\ K_{Ti\min}\le K_{Ti}\le K_{Ti\max},\ j\in S_T \\ P_{Gi\min}\le P_{Gi}\le P_{Gi\max},\ i\in S_G \\ Q_{Gi\min}\le Q_{Gi}\le Q_{Gi\max},\ i\in S_G \\ U_{i\min}\le U_i\le U_{i\max},\ i\in S_D \\ |T_i|\le T_{i\max},\ i\in S_L \end{cases} \tag{11}$$

where vectorvariable $x=[Q_C,K_T,U,\theta,P_G,Q_G]^T$denotes the switching capacity of the reactive power compensation device, the transformertap ratio, the node voltage amplitude, the node voltage phase angle, active and reactive power of generator , respectively.$P_{Di}$and Q_{Di} are the active and reactive load of node i, respectively; $B_{ij}$is the susceptance of line$i$-j; S_C, S_T, S_G, $S_D$and S_L are the set of reactive power compensation devices,transformers, generators, load buses, and lines, respectively.

3. TQL with knowledge transfer

3.1. Knowledge matrix

The Q-value matrix of Q-learning is adopted as the knowledge matrix of TQL and defined as a record of algorithm optimization strategy[14][15][16].The element of knowledge matrix, i.e., $Q(s,a)$, denotes the expected accumulative reward by selecting an action a in a state s.In the process of optimization, the algorithm achieves the convergence after a large number of iterative trial and error, and the process, the optimization body maps the state to the action, is stored in theknowledge matrix.As shown in Figure 1, an agent (here it means a tribal member) can obtain an action policy under a given state from the knowledge matrix and update its prior knowledge matrix by feedback.

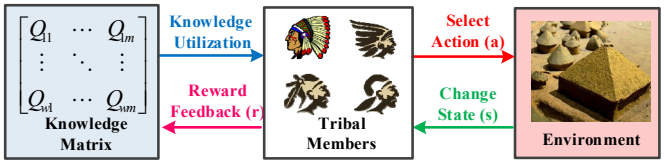

Figure 1.Knowledge matrix.

Basically, the Q-value matrix is a lookup table with the size of $|\boldsymbol{S}|\times|\boldsymbol{A}|$. For large-scale complex systems, the action space$|\boldsymbol{A}|$will increase exponentially with the increase in the number of variables, that is, "Curse of Dimensionality", which results in the calculation difficult to carry out.Therefore, in order to considerably reduce the dimension, the original knowledge matrix is divided into several interrelated low-dimensional sub-matrices. The sub-matrices correspond to the corresponding variables, and the rows and columns of the matrix correspond to the state and action of the variables respectively, and the number of rows of $\boldsymbol{Q}_{i+1}$ is the same as the number of $\boldsymbol{Q}_i$columns.In other words, the action space $\boldsymbol{A}_i$ of the ith variable is also the state space $\boldsymbol{S}_{i+1}$ of the (i+1)th variable.The action selection process of the different variables is no longer isolated, but presents a chain state - action pairs to extend, i.e., the state space $\boldsymbol{S}_{i+1}$ of the (i+1)th variable cannot be selected until the ith variable has been determined.In the knowledge matrix, the elements not only reflect the merits and demerits of the current strategy, but also reflect the compactnessdegree between adjacent variables. The larger the element value, the closer the state-action combination of adjacent variables is.

A multi-agent cooperative mechanism is introduced into TQL algorithm to update the knowledge matrix. When the agents (tribal members) complete a trial and error each time, the algorithm will

assess its fitness and give this selected state - action pair a reward. Then agents update the knowledge matrix according to the given reward value, thus multiple knowledge elements is updated through an iteration.

Compared with the traditional Q-learning algorithm,the matrix convergence process is obviously accelerated. After introducing multi-agent coordination, the knowledge matrix is updated as follows [17]:

$$\rho_{ij}^{k} = R(s_{ij}^{k}, s_{ij}^{k+1}, a_{ij}^{k}) + \gamma \max_{a_i \in A_i} \boldsymbol{Q}_i^{k}(s_i^{k+1}, a) - \boldsymbol{Q}_i^{k}(s_{ij}^{k}, a_{ij}^{k}) \tag{12}$$

$$\boldsymbol{Q}_i^{k+1}(s_{ij}^{k}, a_{ij}^{k}) = \boldsymbol{Q}_i^{k+1}(s_{ij}^{k}, a^{k}) + \alpha \rho_{ij}^{k} s^{k} \tag{13}$$

where*i* denotes the *i*th variable and *j* denotes the *j*thtribal member; α and γ are the learning factor and discount factor, respectively; $R(s_{ij}^{k},s_{ij}^{k+1},a_{ij}^{k})$ is the reward of a transition from state s_{ij}^{k} to state s_{ij}^{k+1} under a selected action a_{ij}^{k} in the *k*th iteration;

3.2. Optimization mode

In order to improve the ability to adapt to the environment, primitive people living in the same habitat will spontaneously form tribes where different tribal members achieve survival and developmentthrough mutual cooperation. Inspired by this primitive human social activity, TQL is able to achieve global convergence and accurate local search through the mutual cooperation by different tribal members.

The algorithm classifies and divides the tribe members according to the reward function values. The top 25% of the reward function are the tribal patriarchs, where the best individual is the chiefs, 25% of the individuals in the middle are civilians, and 50% of the rears are rangers. There are two tendencies, i.e., search and utilization in the optimization mode of reinforcement learning. Focusing on search can enhance its global convergence, and focusing onutilization can improve the convergence rate.

Following the behavioural strategy or chaos search strategies are taken to search by chiefs, tribal patriarchs and civilians, who assume the main search task. Chaos phenomenon is random, regular and ergodic, therefore, this search is conducive to enrich the diversity of the population and jump out of the local optimal solution.

The fitness of tribal patriarchs in the tribe is in a dominant position, which guides civilians and rangers, but follows the chiefs of the lead. Thus, the patriarchstake chaotic searches based on Logistic mapping as the main mode of movement andhave certain following behavior to chiefs. The movement of the patriarch can be described as follows:

$$r_i^{t} = \mu r_i^{t-1}(1 - r_i^{t-1}),\ f_1 = r_i^{t} r_{\text{sign}} step_i^{k} \tag{14}$$

$$f_2 = hr_{\text{sign}}(x_{\text{lead}}^{k} - x_i^{k}) \tag{15}$$

$$x_i^{k} = x_i^{k} + f_1 + f_2 \tag{16}$$

where f_1, f_2 are the chaotic search components and follow the chiefs components, respectively;$\mu$is the chaos control parameter, 4 in this paper; $r_i^{t}$is the random number generated by the chaotic sequence for the *t*th cycle;$r_{\text{sign}}$is a random number with a value of 1 or -1; *h*isthe approximation factor, characterizing the degree of individual followschiefs,this paper takes 0.1;x_{lead} is the chiefs; Step vector$step_i$=($step_i^{1}$， $step_i^{2}$,…， $step_i^{Dim}$),

$$step_i^{k} = \left| x_i^{k} - x_{\text{rand}}^{k} \right|, k = 1, 2, ..., Dim \tag{17}$$

wherex_{rand} is a randomly selected patriarch differ from itself.

Chiefs take the same mode of movement as patriarchs and are also carried out according to Eqs. (14) to (16). The difference is that the follow components toward itself are zero.

The movement of civilians consists of the following components of the chiefs and the patriarchs

$$x_i^{k} = x_i^{k} + c_1 r_1 (x_{\text{lead}}^{k} - x_i^{k}) + c_2 r_2 (x_{\text{str}}^{k} - x_i^{k}) \tag{18}$$

where$D$ is a dimension of the solution component; c_1， c_2 are the following factors of the chiefs and patriarchs, this paper takes 1.5 and 1, respectively;r_1, $r_2$are random numbersamong [0,1], respectively;x_{str} is the closest to i patriarch.

Rangers adopt the Pm-greedy strategy. Under the guidance of the knowledge matrix, rangerssearch in the feasible domain to improve the efficiency of the algorithm, through the utilization of information. It can be described as follows:

$$a_{ij}^{k+1}=\begin{cases}\arg\max\limits_{a^i\in A_i} \boldsymbol{Q}_i^{k+1}(s_{ij}^{k+1},a_i) & r_3\geq P_{\mathrm{m}}\\ a_s & r_3<P_m\end{cases} \tag{19}$$

where$r_3\in[0,1]$ is the random number;$P_{\mathrm{m}}\in[0,1]$ denotes migration probability; $a_s$means roulette selection. When $r_3< P_{\mathrm{m}}$, the ranger chooses roulette according to the action probability matrix $\boldsymbol{P}_i$; when $r_3\geq P_{\mathrm{m}}$, the ranger chooses the action that is expected to accumulate the maximum reward in the current state, that is, the implementation of the greedy strategy.

The action probability matrix $\boldsymbol{P}_i$ denotes the probability of selection of state-action pair,and has a positive correlation with the value of the knowledge matrix element $\boldsymbol{Q}_i(s_i,a_i)$. Pi is updated as follows:

$$\begin{cases}\boldsymbol{e}_i(s_i,a_i)=\dfrac{1}{\boldsymbol{Q}_i(s_i,a_i)-\beta\max\limits_{a'\in A_i}\boldsymbol{Q}_i(s_i,a')}\\ \boldsymbol{P}_i(s_i,a_i)=\dfrac{\boldsymbol{e}_i(s_i,a_i)}{\sum\limits_{a'\in A_i}\boldsymbol{e}_i(s_i,a')}\end{cases} \tag{20}$$

where$\beta$is the divergence factor to magnify the divergence of sub-matrices and e^i is the transition matrix.

After each round of iteration, all the tribal members have completed the search and got the feedback of the current round of reward function. According to the reward function, the tribal members are reordered and assigned new roles. Ranking front tribal members become chiefs, patriarchs and civilians, sorted by the later became rangers, where the former chiefs to maintain the original position unchanged. Therefore, the algorithm not only guarantees the continuity of the elite individual, but also maintains the global search performance in the solution space.

3.3. Transfer learning

If the task ready to be completed by TQL contains multiple similar tasks, then the efficiency of new tasks can be improved through knowledge transfer.

As illustrated in Figure 2, based on the existing knowledge of source task, knowledge transfer can accelerate the learning process of new tasks. In order to acquire the initial knowledge of similar new tasks, the source task must be studied in pre-learning at first. Assuming that the action space and the state space keep constant, the optimal knowledge matrices of the source tasks can be treated as the initial knowledge matrices of the target tasks. In the transfer learning process, the optimal knowledge matrices of source tasks $\boldsymbol{Q}_{\mathrm{S}}$ can be transferred to the initial knowledge matrices of similar new tasks $\boldsymbol{Q}_{\mathrm{N}}$.

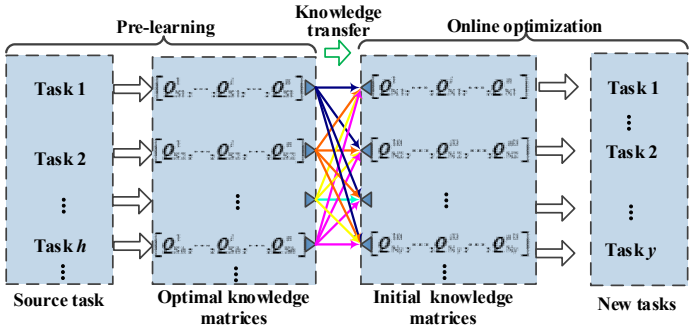

Figure 2.The procedure of knowledge transfer.

The optimal knowledge matrix of the source task contains both related and invalidinformation about new tasks. Therefore, once the related knowledge cannot be fully extracted, invalid information will interfere with the process of new task learning, which will reduce the effect of transfer learning, i.e., malignant negative transfer. To handle this, new tasks only extract the closely relevant knowledge and takes similarity as the criterion to select the object for learning during the process of transfer learning

In RBRPO, a task corresponds to a time section, and the demand for active load at different time scenarios is different. Since the solution of RBRPO is mainly determined by the power flow of the system, the active power deviation of different time scenarios is defined as the similarity between source and new tasks.

Assuming that the active power demand of the new task *x*is P_{Dx}, the two source tasks with the least active power deviations from task *x* are task *i* and task *k*, and $P_{Di}<P_{Dx}<P_{Dk}$ is satisfied, the similarity between task *x* and the two source tasks can be calculated as follows:

$$\begin{cases} \eta_1 = \dfrac{P_{Dx} - P_{Dj}}{P_{Dk} - P_{Dj}} \\ \eta_2 = \dfrac{P_{Dk} - P_{Dx}}{P_{Dk} - P_{Dj}} \end{cases} \tag{21}$$

whereη_1 and η_2 are the similarities weighting factors, with $\eta_1+\eta_2=1$.

The knowledge matrix of the new task *x* can be obtained by a linear transfer, which yields

$$\boldsymbol{Q}_x^i = \eta_1 \boldsymbol{Q}_j^i + \eta_2 \boldsymbol{Q}_k^i \tag{22}$$

where$\boldsymbol{Q}_x^i$,$\boldsymbol{Q}_j^i$and $\boldsymbol{Q}_k^i$ denote the knowledge sub-matrices of the *i*th variable in source task *x*, source task *j* and new task *k*, respectively.

4. Case Studies

In this paper, the TQL algorithm is used to solve RBRPOon the IEEE 118-bus system, which performance is compared with that of GA[5], PSO[6], ABC [7], ANT-Q[18], quantum genetic algorithm (QGA) [19], ant colony optimization(ACO) [20].Simulation is undertaken in MatlabR2014a by a personal computer with Intel(R) Core TM i7-6700 CPU at 3.40GHz with 16GB of RAM. The power flow calculation is based on the Matpower6.0toolbox in Matlab R2014a.

4.1. Simulation Model

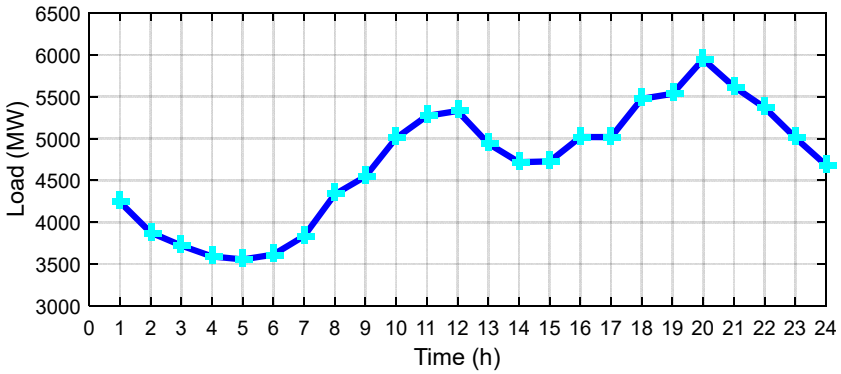

Figure 3.A typical daily load curve.

IEEE 118-bussystem consists of 54 generators and 186 branches, which is divided into three voltage levels, i.e.138kV, 161kV, 345kV. The number of controllablevariables of IEEE 118-bussystem is 25, which contain the reactive power compensation capacity of the shunt capacitor, the ratio of the on-load tap changer and the terminal voltage of the generator. More specifically, the compensation capacity of the shunt capacitor is divided into [-40%, -20%, 0%, 20%, 40%]from its

nominal level, transformer ratio is divided into three grades as [0.98, 1.00, 1.02] and the generator terminal voltage is uniformly divided into seven grades as [1.00, 1.01, 1.02, 1.03, 1.04, 1.05, 1.06].

The figure 3 shows the typical load curve of the IEEE 118-bus system. Based on the demand of active power, load can be uniformly divided into 7 intervals,{[3556, 3897), [3897, 4239), …, [5604, 5945]}. Therefore, the number of source tasks for the IEEE 118-bus system is 8.

4.2. Algorithm comparison

The optimization of the objective functionsperformed by each algorithm within 24 hours of the day is shown infigure 4, where the blue solid line represents the optimization result of the TQL, and the dotted line represents the other algorithm. It can be seen from the figure that the trend of the objective function of TQL is basically the same as other algorithms in one day, and its objective function curve is only slightly higher than ACO algorithm and superior to other AI algorithm. It shows that the algorithm can take full advantage of the related knowledge obtained in the pre-learning process and avoid the occurrence of the negative transfer, and has good global convergence performance by fully grasping the similarity between the source task and the new tasks.

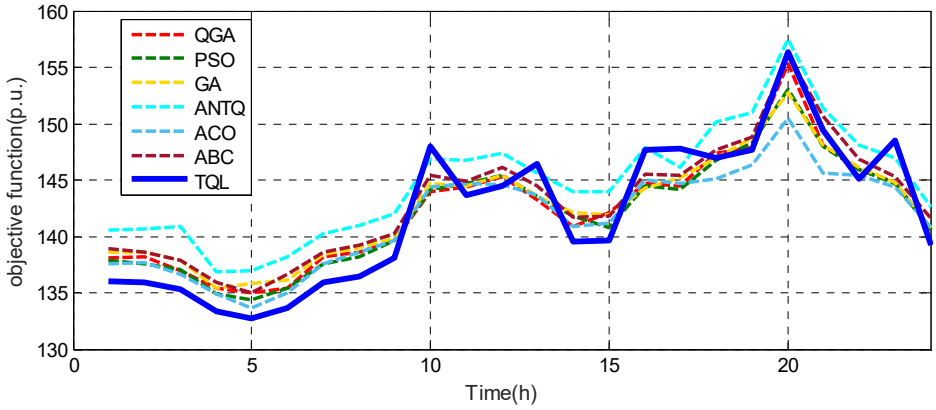

Figure 4.Optimization results of objective function of 24section obtained by each algorithm.

Ingeneral, the solving process of AI algorithm is random and uncertain. In order to compare the optimal performance of each algorithm correctly, each algorithm runs 10 times. Since the algorithms carry outeach 24-hour load level optimization in each round of simulation, for each algorithm, the total number of simulation is10×24=240times, the convergence of the algorithm has been fully reflected. Table 1 indicates the average data of the objective functions obtained from the 10 runs of each algorithm. The values of the power loss, the voltage deviation component, the risk index and the objective function are the sum of 24 tasks. The calculation time is the sum of time for each algorithm to complete 24 tasks. The convergence time is the average time to complete a single task. It should be noted that the convergence effect of the algorithm is only determined by the objective function value instead ofthe power loss, the voltage stable component or the risk index.

It can be seen from the table1, TQL algorithm only needs about 895s to complete the optimization of 24 tasks,whichis much faster than the other algorithms. Moreover, the convergence rate of TQL is 2~20 times faster than the other 6 algorithms, averaging more than 10 times that of other algorithms. The objective function value obtained by TQL is 3407.87, which places second among all 7 algorithms. However, its optimization performance of power loss and the voltage deviation component are better than ACO algorithm. This shows that TQL fully exploits the similarities between source and new tasks, and significantly accelerates the optimization process through knowledge transfer. At the same time, TQL combines the trial and error iteration mechanism of Q-learning with the random optimization mode of tribe organically to ensure the global convergence performance of the algorithm. The figure 5 and figure6 show the speed advantage and excellent search ability of TQL algorithm intuitively.

Table 1.Simulation Results of Each Algorithms on IEEE 118-bus System in 10 Times.

Algorithm	Calculation time (s)	Convergence time (s)	P_{loss}(MW)	V_d(%)	Risk index (p.u.)	Objective function (p.u.)
ABC	13207.65	550.31	2796.97	378.07	0.0258	3433.43
ANT-Q	6183.42	255.76	2809.02	391.63	0.0272	3473.66
PSO	17365.14	723.54	2795.72	371.65	0.0266	3408.83
GA	1806.11	75.25	2802.30	379.16	0.0238	3419.48
QGA	13090.59	545.44	2784.28	377.76	0.0252	3414.18
ACO	8575.11	357.29	2790.39	375.17	**0.0232**	**3398.39**
TQL	**894.85**	**37.28**	**2774.39**	**369.06**	0.0264	3407.87

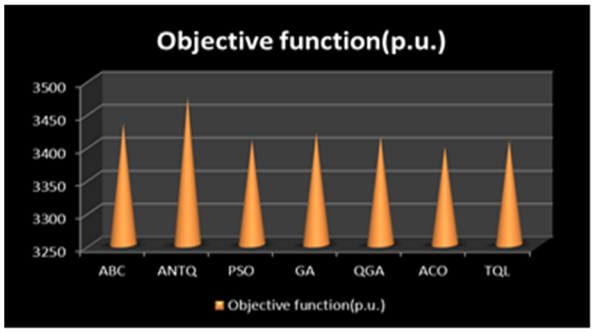

Figure 5.The averageobjective function of IEEE 118-bus system obtained by different algorithms in 10 times.

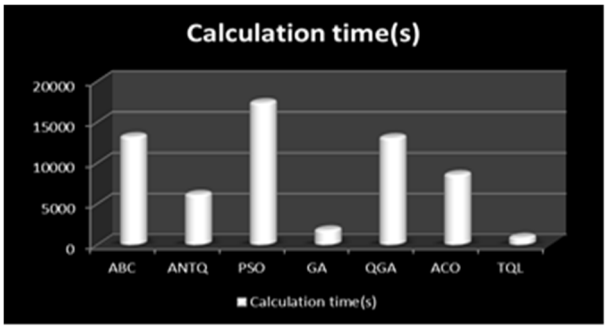

Figure 6. The totalcalculation time of IEEE 118-bus system consumed by different algorithms in 10 times.

The objective function convergence performance statistics in 10 simulations of each algorithm are shown in table 2. In the table, variance, standard deviation and relative standard deviation are calculated to evaluate the convergence stability of the algorithm. TQL shows the best performance among all AI algorithm, especially its convergence stability, where its relative standard deviation is only 40% of the ANT-Q algorithm. This is because TQL reduces the global search blindness and uncertainty through transferor using the past knowledge.

Table 2.Statistical Results of Objective Function of IEEE 118-bus System in 10 Times.

Algorithm	Worst	Best	Variance	Standard Deviation	Relative Standard Deviation
ABC	3426.09	3440.19	17.40	4.17	1.28E-03
ANT-Q	3458.39	3483.53	53.71	7.32	2.22E-03
PSO	3403.34	3414.07	16.36	4.04	1.25E-03
GA	3411.07	3430.18	36.67	6.05	1.86E-03
QGA	3409.60	3420.62	11.88	3.44	1.06E-03
ACO	**3391.23**	**3401.90**	10.44	3.23	0.95E-03
TQL	3403.30	3411.35	**8.17**	**2.85**	**0.88E-03**

4.3. Optimization analysis

The node voltage and the power flow distribution of the IEEE 118-bus system at load section 20 are shown in the figure 7 and figure 8, respectively, which compare the situation before and after TQL optimization. After the optimization, the voltage deviation of the system node is reduced and distributed in the range of [0.96,1.04]. Therefore, the risk of the system node voltage overrun has been effectively controlled.It can be seen from the figure 8 that after the optimization, the distribution of the power flow is more uniform, which avoids the occurrence of branch overload.

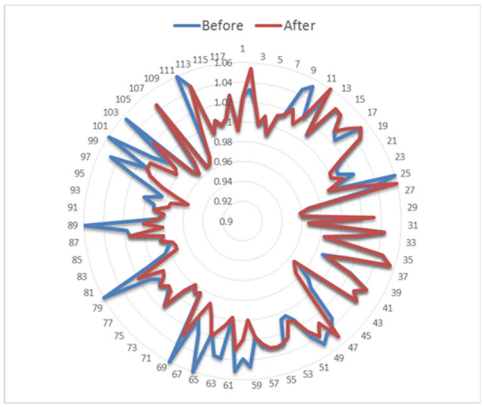

Figure 7.The node voltage distribution of the IEEE 118-bus system at load section 20.

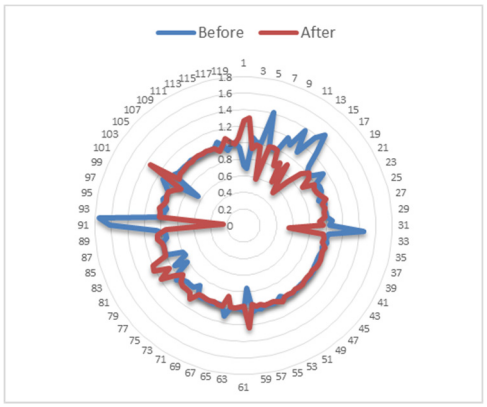

Figure 8.The power flow distribution of the IEEE 118-bus system at load section 20.

The distribution radar map of the objective function and sub-objective are shown in figure 9, which compares the different results before and after optimization. After the optimization of TQL, the values of the power loss, voltage deviation, risk index and objective function are all less than the values before optimization, which verifies the validity of the multi-objective reactive power optimization model proposed in this paper. Among them, the improvement of system operation risk is the most obvious, the optimized risk index is only 68% of its values before optimization, which means that the ability to resist the power system uncertainty risk has been significantly improved through optimization by TQL.

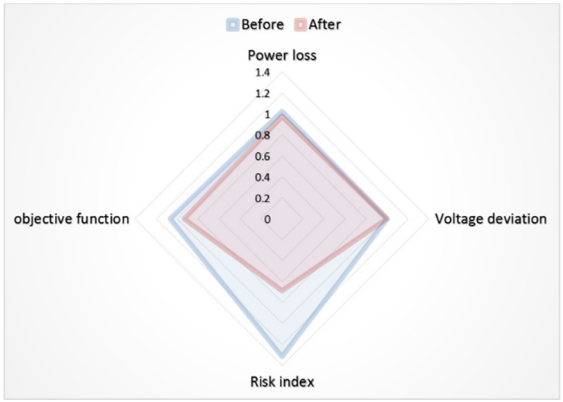

Figure 9.The distribution radar map of the objective function and sub-objective.

5. Conclusion

In this paper, a novel TQL algorithm is proposed for RBRPO. The main innovations can be summarized as follows:

1) The theory of risk assessment is introduced into the tradition reactivepower optimization issue and the RBRPO model is proposed. The model aims at reducing the operating risk and the power loss of the system and optimizing the system voltage stability at the same time, which is beneficial to the security and economical operation of the power system.

2) TQL combines the trial and error iteration mechanism of Q-learning with the random optimization mode of tribeorganically to ensure the local depth search ability and global convergence performance of the algorithm.

3) The active load deviation is defined as the similarity degree. TQL can accelerate the optimization process of the new task by knowledge transfer using knowledge of similar source tasks.

The convergence stability and excellent performance of TQL can be confirmed by the simulation results of IEEE 118-bus system, i.e., from 2 to 20 times faster than that of existing AI algorithms for RBRPO, while the quality of optimal solution and the convergence stability can be guaranteed. Thus

TQL can be a useful tool to deal with the risk-based reactive power optimization issue in power system.

References

[1] Grudinin N. "Reactive power optimization using successive quadratic programming method." *Power Systems IEEE Transactions on***13.4**(1998):1219-1225.
[2] Chen Y L, and Liu C C. "Interactive fuzzy satisfying method for optimal multi-objective VAr planning in power systems."*IEE Proceedings - Generation, Transmission and Distribution***141.6(1994)**:554-560.
[3] Singh C, Luo X, and Kim H. "Power System Adequacy and Security Calculations Using Monte Carlo Simulation incorporating Intelligent System Methodology." *International Conference on Probabilistic Methods Applied To Power Systems* IEEE,2006, pp:1-9.
[4] Dai J F, Zhou S X, Zong-Xiang L U, Zhu L Z and Zhi Z. "Study on Reactive Power Optimization Based Upon Risk." *Proceedings of the Csee*(2007).
[5] Iba K. "Reactive power optimization by genetic algorithm." *IEEE Transactions on Power Systems***9.2**(1994):685-692.
[6] Zhang W and Liu Y. "Reactive power optimization based on PSO in a practical power system." *Power Engineering Society General Meeting* IEEE,**1**(2004):239-243.
[7] Ozturk A, Cobanli S, Erdogmus P and Tosun S. "Reactive power optimization with artificial bee colony algorithm." *Scientific Research & Essays***5.5**(2010):2848-2857.
[8] He S, Wu Q H and Saunders J R. "Group search optimizer: an optimization algorithm inspired by animal searching behavior." *IEEE Transactions on Evolutionary Computation* **13.5(2009)**:973-990.
[9] Gomez J F,Khodr H M, De Oliveira P M andOcque L. "Ant colony system algorithm for the planning of primary distribution circuits." *Power Systems IEEE Transactions on* **19.2**(2004):996-1004.
[10] Dinh T T H, Chu T H and Nguyen Q U. "Transfer learning in Genetic Programming."*Evolutionary Computation* IEEE, 2015, pp:1145-1151.
[11] Ni M,Mccalley J D, Vittal V andTayyib T. "Online Risk-Based Security Assessment." *IEEE Power Engineering Review***22.11**(2003):59-59.
[12] Li W. *Risk Assessment of Power Systems: Models, Methods, and Applications.*Wiley-IEEE Press, 2005.
[13] Zhang X, Xu H, Yu T, Yang B and Xu M. "Robust collaborative consensus algorithm for decentralized economic dispatch with a practical communication network." *Electric Power Systems Research***140**(2016):597-610.
[14] Zhang X, Yu T, Yang B and Cheng L. "Accelerating bio-inspired optimizer with transfer reinforcement learning for reactive power optimization." *Knowledge-Based Systems* **116**(2016).
[15] Pan S J and Yang Q. A Survey on Transfer Learning.*IEEE Transactions on Knowledge & Data Engineering*, **22.10**(2010):1345-1359.
[16] Taylor M E and Stone P. "Transfer Learning for Reinforcement Learning Domains: A Survey." *Journal of Machine Learning Research*,**10.10**(2009):1633-1685.
[17] Watkins C J C H and Dayan P. "Q-learning." *Machine Learning* **8**, pp(1992):279--292.
[18] Dorigo M and Gambardella L M. "A Study of Some Properties of Ant-Q." *International Conference on Parallel Problem Solving From Nature*, Springer-Verlag, **1141**(1996):656-665.
[19] Malossini A, Blanzieri E and Calarco T. "Quantum Genetic Optimization." *IEEE Transactions on Evolutionary Computation***12.2**(2007):231-241.
[20] Hong H and Liu F. "Binary Adaptive Ant Colony Optimization in Reactive Power Optimization." *Advanced Materials Research***616-618**(2013):2091-2096.