frontiers in Energy Research

BRIEF RESEARCH REPORT
published: 30 September 2021
doi: 10.3389/fenrg.2021.760525

Intelligent Frequency Control Strategy Based on Reinforcement Learning of Multi-Objective Collaborative Reward Function

*Lei Zhang, Yumiao Xie, Jing Ye *, Tianliang Xue, Jiangzhou Cheng, Zhenhua Li and Tao Zhang*

College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, China

Large scale wind power integration into the power grid will pose a serious threat to the frequency control of power system. If only Control Performance Standard (CPS) index is used as the evaluation standard of frequency quality, it will easily lead to short-term centralized frequency crossing, which will affect the effect of intelligent Automatic Generation Control (AGC) on frequency quality. In order to solve this problem, a multi-objective collaborative reward function is constructed by introducing a collaborative evaluation mechanism with multiple evaluation indexes. In addition, Negotiated W-Learning strategy is proposed to globally optimize the solution of the objective function from multi dimensions, it avoids the poor learning efficiency of the traditional Greedy strategy. The AGC control model simulation of standard two area interconnected power grid shows that the proposed intelligent strategy can effectively improve the frequency control performance and improve the frequency quality of the system in the whole-time scale.

Keywords: wind power grid-connected, intelligent frequency control strategy, multi-dimensional frequency control performance standard, Negotiated W-Learning algorithm, global optimization

OPEN ACCESS

Edited by:
Zhenhao Tang,
Northeast Electric Power University, China

Reviewed by:
Zhu Zhang,
Hefei University of Technology, China
Yuanchao Hu,
Shandong University of Technology, China

****Correspondence:***
Jing Ye
x1620730050@163.com

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *18 August 2021*
Accepted: *14 September 2021*
Published: *30 September 2021*

Citation:
Zhang L, Xie Y, Ye J, Xue T, Cheng J, Li Z and Zhang T (2021) Intelligent Frequency Control Strategy Based on Reinforcement Learning of Multi-Objective Collaborative Reward Function. Front. Energy Res. 9:760525. doi: 10.3389/fenrg.2021.760525

1 INTRODUCTION

Automatic Generation Control (AGC) is an important means to realize the balance of active power-load supply and demand in the power system. Among them, the quality of frequency control strategy is an important factor that affects the performance of AGC control (Alhelou et al., 2018; Shen et al., 2021a; Shen and Raksincharoensak, 2021a). However, the control strategies applied in engineering, such as the threshold zone AGC control strategy that takes into account the combined effects of the proportional component, integral component and Control Performance Standard (CPS) control component of the regional control deviation (Arya and Kumar, 2017; Shen et al., 2020a; Xi et al., 2020; Shen and Raksincharoensak, 2021b), have been unable to adapt to the increasingly complex frequency control of interconnected power grids (Shen et al., 2017; Zhang and Luo, 2018).

In recent years, the intelligent frequency control strategy of reinforcement learning has received lots of attention (Yu et al., 2011; Abouheaf et al., 2019; Xi et al., 2019; Shen et al., 2020b; Liu et al., 2020), because it does not rely on models and does not require precise training samples or system prior knowledge (Watkins and Dayan, 1992; Yang et al., 2018; Li et al., 2020; Yang et al., 2021a; Shen et al., 2021b).

However, most intelligent control strategies are built on the CPS frequency control performance evaluation standard. The CPS index has low sensitivity for short-term inter-area power support

evaluation, and cannot take into account the short-term benefits of frequency control performance (Kumar and Singh, 2019; Yang et al., 2019; Zhu et al., 2019). In a system with large-scale wind power grid connection, the ability of each region to comply with CPS indicators is limited. The intelligent AGC control strategy that only considers the CPS control criteria can easily cause short-term concentrated frequency crossings, which seriously affects the control effect of the intelligent AGC control strategy (Wang and James, 2013; Xie et al., 2017; Yang et al., 2021b).

In fact, with the development of grid-connected new energy sources and smart grids, the grid frequency control evaluation standard is transitioning from single-scale evaluation to multi-time-scale and multi-dimensional evaluation. The North American Electric Reliability Council (NERC) proposed a new frequency evaluation performance index named Balancing Authority ACE Limits (BAAL), which is used to ensure the short-term frequency quality of the system by constraining the mean value of the frequency difference fluctuates in any 30 min not to exceed the limit. However, the intelligent AGC control strategy under both BAAL and CPS indicators is a kind of multi-objective control problem, and there is no relevant literature to study it.

In response to the above problems, this paper proposes an intelligent frequency control strategy for collaborative evaluation of multi-dimensional control standards. This strategy constructs and introduces a collaborative reward function that considers the CPS index and the BAAL index in the multi-objective reinforcement learning algorithm. Then, the Negotiated W-Learning strategy is used to learn the action space of the agent, which effectively solves the problem that the agent cannot fully explore the action (Nathan and Ballard, 2003; Liu et al., 2018; Wang et al., 2019). Simulation examples show that the proposed intelligent control strategy can effectively improve the overall frequency performance quality of the power system.

2 FREQUENCY CONTROL PERFORMANCE EVALUATION STANDARD OF INTERCONNECTED POWER GRID

2.1 CPS1 Frequency Control Performance Evaluation Standard

NERC uses the BAL (BAL-001) disturbance control series of indicators to evaluate the frequency control quality of the interconnected power grid. Among them, the CPS1 (BAL-001-2: R1) indicator is the most widely used in China, as shown in **Eq. 1**:

$$AVG_{1,T}\left[\left(\frac{ACE^{m}_{1\,\mathrm{min}}}{-10B_m}\cdot\Delta F_{1\,\mathrm{min}}\right)\right]\le\varepsilon^2 \tag{1}$$

where $\Delta F_{1\ \mathrm{min}}$ and $ACE^{m}_{1\ \mathrm{min}}$ are separately the average value of the frequency deviation and power deviation in the control area within 1 min, B_m is the frequency deviation coefficient of the area m, and represents the frequency adjustment responsibility assigned to area m. $AVG_{1,T}$ (·) means calculate the average value for 12 months, ε is the upper limit of the area m in controlling the frequency deviation.

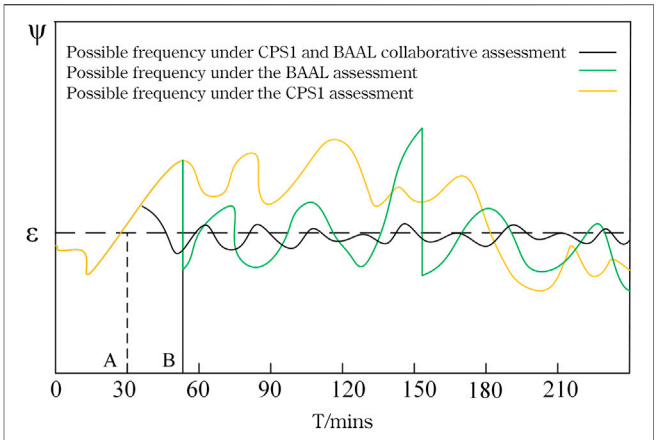

FIGURE 1 | The distribution curve of the integrated frequency deviation factor over time.

Taking the situation that the actual frequency is higher than the planned frequency as an example, expand **Eq. 1** as follows:

$$\frac{1}{T}\int_0^T\frac{\Delta F}{\varepsilon}*\left[\frac{\Delta P_{\mathrm{tie}}}{-10B_m\varepsilon}+\frac{\Delta F}{\varepsilon}\right]dt\le 1 \tag{2}$$

where: T is the entire time period, $\Delta F/\varepsilon$ is the frequency deviation contribution of the region itself, $\Delta P_{\mathrm{tie}}/-10B_m\varepsilon$ is the frequency contribution of other regions to this region, and $\Delta P_{\mathrm{tie}}/-10\ \mathrm{B}_m\varepsilon+\Delta F/\varepsilon$ is the comprehensive frequency deviation contribution. For the convenience of analysis, define $\Delta F/\varepsilon*[\Delta P_{\mathrm{tie}}/-10\ \mathrm{B}_m\varepsilon+\Delta F/\varepsilon]$ as the comprehensive frequency deviation factor, and denoted by ψ.

The CPS1 indicator statistically evaluates the rolling root mean square of the frequency difference time series during the T period in the evaluation area. When T is large enough, the system frequency deviation qualification rate is greater than 99.99%. Therefore, CPS1 is a long-term evaluation index reflecting the frequency quality of interconnected power grids.

2.2 BAAL Frequency Control Performance Evaluation Standard

NERC proposed the BAAL (BAL-001-2: R2) evaluation index in 2013 and began to implement it in 2016, as shown in **Eq. 3** ~4:

$$T\left[ACE^{m}_{1\,\mathrm{min}}\ge-10B_m\frac{\left(F_{FIL-high}-F_s\right)^2}{\left(F_A-F_S\right)_{1\,\mathrm{min}}}\right]\le T_v \tag{3}$$

$$T\left[ACE^{m}_{1\,\mathrm{min}}\le 10B_m\frac{\left(F_{FIL-low}-F_s\right)^2}{\left(F_A-F_S\right)_{1\,\mathrm{min}}}\right]\le T_v \tag{4}$$

where F_A is the actual frequency value; F_S is the planned frequency value; $F_{\mathrm{FTL\text{-}high}}/F_{\mathrm{FTL\text{-}low}}$ is the high/low frequency trigger limit; $\mathrm{T_v}$ is the specified allowable continuous time limit. T [·] is the continuous over-limit time.

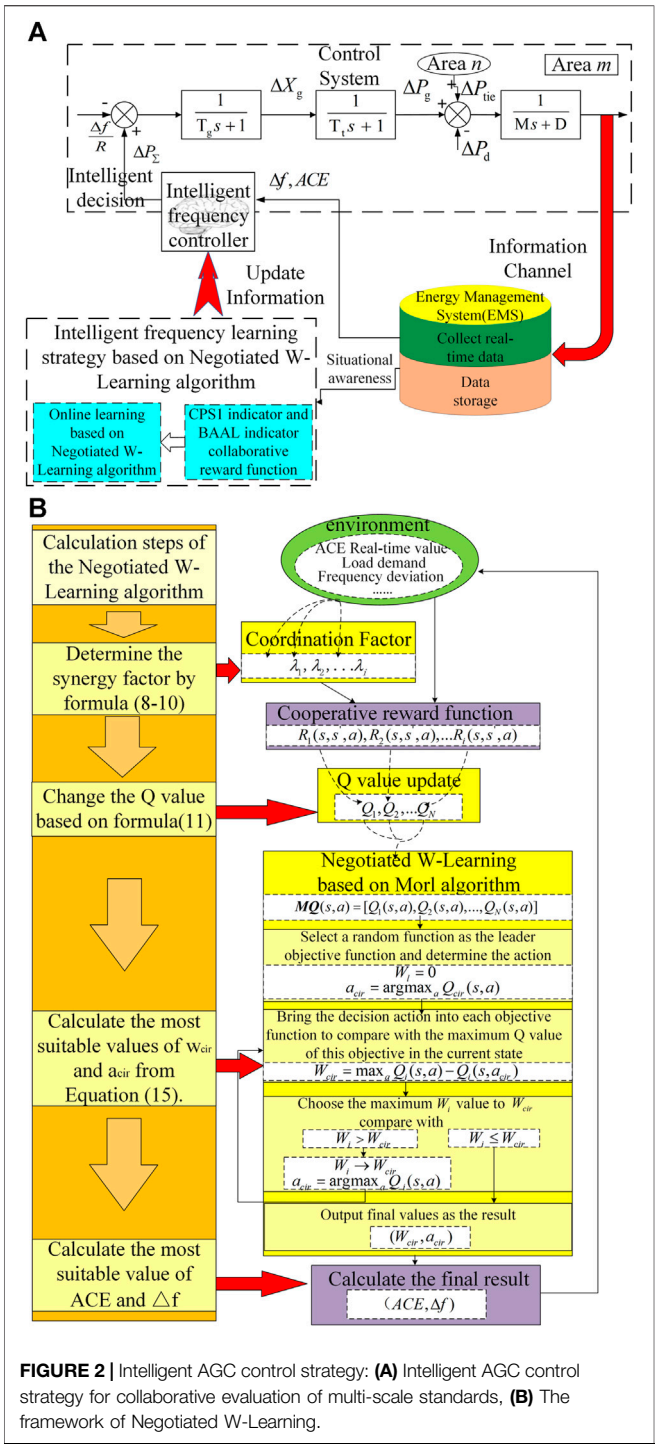

FIGURE 2 | Intelligent AGC control strategy: **(A)** Intelligent AGC control strategy for collaborative evaluation of multi-scale standards, **(B)** The framework of Negotiated W-Learning.

Taking the situation that the actual frequency is higher than the planned frequency as an example, **Eq. 3** can be transformed into the following form in the same way:

$$T\left[\frac{1}{T^n}\int_{T'}^{T'+T''}\frac{\Delta F}{\varepsilon}^{*}\left[\frac{\Delta P_{\text{tie}}}{-10\ \text{B}_m\varepsilon}+\frac{\Delta F}{\varepsilon}\right]dt\geq 1\right]\leq \text{T}_\text{v} \qquad (5)$$

2.3 Performance Analysis Under the Joint Control of BAAL Standard and CPS1 Standard

In order to further study the feature of the two index, **Figure 1** shows the change curve of the comprehensive frequency deviation factor ψ, which considers different performance indicators under the influence of the time dimension.

As shown in **Figure 1**, taking point A as the critical point of frequency line crossing, when only CPS1 is considered, the system frequency can still meet the requirements of control performance index, but it will affect the safe operation of various equipment in the system and cause the power quality reduced. If only the BAAL indicator is considered, the system frequency may appear "vertical dro" and "tip oscillatio," as shown in point B in **Figure 1**. At this time, the synchronous generator frequently receives the opposite frequency deviation signal that occurs in a short period of time. This situation will increase the wear of the unit. When considering the effects of CPS1 and BAAL indicators at the same time, the frequency will change into the reverse process under the influence of BAAL performance after short-term limit violation.

In summary, if CPS1 and BAAL indicators can be coordinated to constrain the system frequency closely, it can guarantee not only the long-term frequency quality but also the short-term frequency safety.

3 INTELLIGENT AGC CONTROL STRATEGY CONSIDERING COOPERATIVE EVALUATION OF MULTI-DIMENSIONAL CONTROL STANDARDS

Based on the analysis in **Section 2.3**, this paper constructs an AGC control model based on a multi-objective collaborative reward function reinforcement learning frequency control strategy. As shown in **Figure 2A**, it mainly consists of the following parts: system governor, equivalent module of the generator, dynamic model of system's frequency deviation, and intelligent brain controller. Where R, T_g, T_t, M, D are separately the equivalent unit adjustment coefficient, time constant of the governor, equivalent generator time constant, equivalent inertia coefficient and equivalent damping coefficient of the power system in area m; ΔP_{tie} is the exchange power deviation of the tie line in area m, ΔX_g, ΔP_g, ΔP_d are separately the change in the position of the regulating valve, in generator output power and in load disturbance, ΔP_Σ is the total adjustment command of the unit.

Frequency controller intelligent learning stage: This article uses a multi-objective collaborative reward function reinforcement learning strategy to learn and train the intelligent frequency controller. This strategy mainly includes two parts, namely CPS1 index and BAAL index cooperative reward function and Negotiated W-Learning based intelligent frequency control learning algorithm. First, use the MORL idea to construct the instant reward function of CPS1 index and BAAL

index, and use dynamic coordination factors to characterize the impact of different indicators on environmental changes. Then, the implementation rewards given under the MORL learning are used to update the respective state action sets of the CPS1 index and the BAAL index. Finally, Negotiated W-Learning conducts a global search to get the final action, which will meet the CPS1 and BAAL indicators and environmental feedback characteristic information.

Frequency controller online deployment stage: The learned and mature frequency controller receives the SCADA database in the Energy Management System (EMS) in each AGC control cycle to collect frequency deviation, ACE, CPS, BAAL, and other data in real time, and make real-time frequency control action.

3.1 Collaborative Reward Function of CPS1 Indicator and BAAL Indicator

This paper constructs a cooperative reward function based on the CPS1 indicator and the BAAL indicator, which is expressed as follows:

$$\begin{aligned} R_1(s,s',a) &= -\lambda_1 (ACE - BAAL)^2 \\ R_2(s,s',a) &= -\lambda_2 \left(CPS1^* - CPS1\right)^2. \end{aligned} \tag{6}$$

Among them: $R_i(s,s',a)$ is the instant reward value obtained when the ith goal is transferred from state s to state s' through action a; ACE (t) is the real-time value of the regional control deviation at the current moment; s is the system state $[ACE(t)]$ at time t, s' is the state $[ACE\ (t+1)]$ at time $t+1$, a is the system action $(\Delta P_{\Sigma}(t))$ when the system goes from s to state s'. $BAAL(t)$ is the instantaneous value of BAAL at time t, $CPS1$ (t) is the instantaneous value of $CPS1$ at time t, $CPS1^*$ is the target value, generally 200%.

λ_i is the dynamic coordination factor of the cooperative reward function, that is, λ_i changes dynamically with each state transition process. This paper adopts the method of comprehensive weighting and multiplicative weighting, comprehensively considers the preferences of decision makers and the inherent statistical law between the index data to determine the value of the dynamic coordination factor.

Firstly, Define parameter K as a parameter for evaluating the importance of frequency performance evaluation indicators. $K_{i,j}$ represents the importance degree of the evaluation index relative to another one in the frequency performance evaluation. When there is an out-of-bounds situation such as ACE < BAAL or $CPS1$ > 200, the importance of the corresponding indicators will increase accordingly. When the two indicators play equal or unimportant roles in the frequency evaluation process, the corresponding $K_{i,j}/K_{j,i}$ values are all 4 or 0. The relative importance of any index increases by one point, the corresponding $K_{i,j}/K_{j,i}$ value increases by 1, and the $K_{j,i}/K_{i,j}$ value decreases by 1. Then obtain the weighting factors of each target in each action cycle:

$$w_i = \frac{K_{j,i}}{K_{j,i} + K_{i,j}} \ (i \neq j) \tag{7}$$

In order to eliminate subjectivity, the entropy method is used to calculate the coefficient of difference between the two indicators β_i:

$$\beta_i = \frac{1 + \ln^{-1}(N) \sum_{y=1}^{K} P_{y,i} \ln\left(P_{y,i}\right)}{\sum_{i=1}^{N}\left(1 + \ln^{-1}(N) \sum_{y=1}^{K} P_{y,i} \ln\left(P_{y,i}\right)\right)} \tag{8}$$

$$P_{y,i} = x_{y,i} \Big/ \sum_{y=1}^{K} x_{y,i} \tag{9}$$

Where: $x_{y,i}$ is the standardized index value of the ith frequency control performance evaluation index at the yth time, K represents the number of the ith frequency control performance evaluation index from 0 to the current time t, and N represents the target number. $P_{y,i}$ is the proportion of $x_{y,i}$ to the total number of indicators from 0 to t.

At last, the final coordination factor is determined by multiplication weighted method. Therefore, the coordination factor can be obtained by combining 8 and 9:

$$\lambda_i = \frac{\sqrt{w_i \beta_i}}{\sum_{i=1}^{N} \sqrt{w_i \beta_i}} \tag{10}$$

3.2 Negotiated W-Learning Intelligent Frequency Control Learning Algorithm

The update formula of MORL is the same as the state-action value function update of traditional Q learning, as shown in **Eq. 11**. In order to facilitate the selection of the optimal action that satisfies each of the following goals, this paper uses the $\boldsymbol{MQ}(s,a)$ vector to represent the state-action value function Q value of the action a in the state s for the N goals, as shown in **Eq. 12**, and the optimal action strategy π^*_{MQ} for each target in the current state expressed in **Eq. 13**:

$$Q_i(s,a) \leftarrow Q_i(s,a) + \alpha\left(R_i(s,s',a) + \gamma \max_{a \in A} Q_i(s',a) - Q_i(s,a)\right) \tag{11}$$

$$\boldsymbol{MQ}(s,a) = [Q_1(s,a), Q_2(s,a), \ldots, Q_N(s,a)] \tag{12}$$

$$\pi^*_{MQ} = \arg\max_{a}\left\{\max_{i} \boldsymbol{MQ}(s,a)\right\} \tag{13}$$

In **Eq. 11**:α $(0 < \alpha < 1)$ is the learning rate, which is set to 0.01 in this article; γ is the discount coefficient, which is set to 0.9 in this article; Q_i (s, a) represents the Q value of the ith target's choice of action a in state s.

However, the above-mentioned optimal action selection strategy cannot guarantee that the agent fully explores the entire state-action space. In this paper, Negotiated W-learning strategy is used to optimize the $\boldsymbol{MQ}$ (s, a) vector space. This strategy defines variable W_i as a leader parameter. The operation steps are as follows, and **Figure 2B** is a reference flow chart:

Step 1: Choose an objective function in the $\boldsymbol{MQ}(s, a)$ vector space as the guide objective function. Its investigation parameter is expressed as W_i. The first guide objective function is uniformly set to $W_{cir} = 0$, and the guide action is obtained as follows:

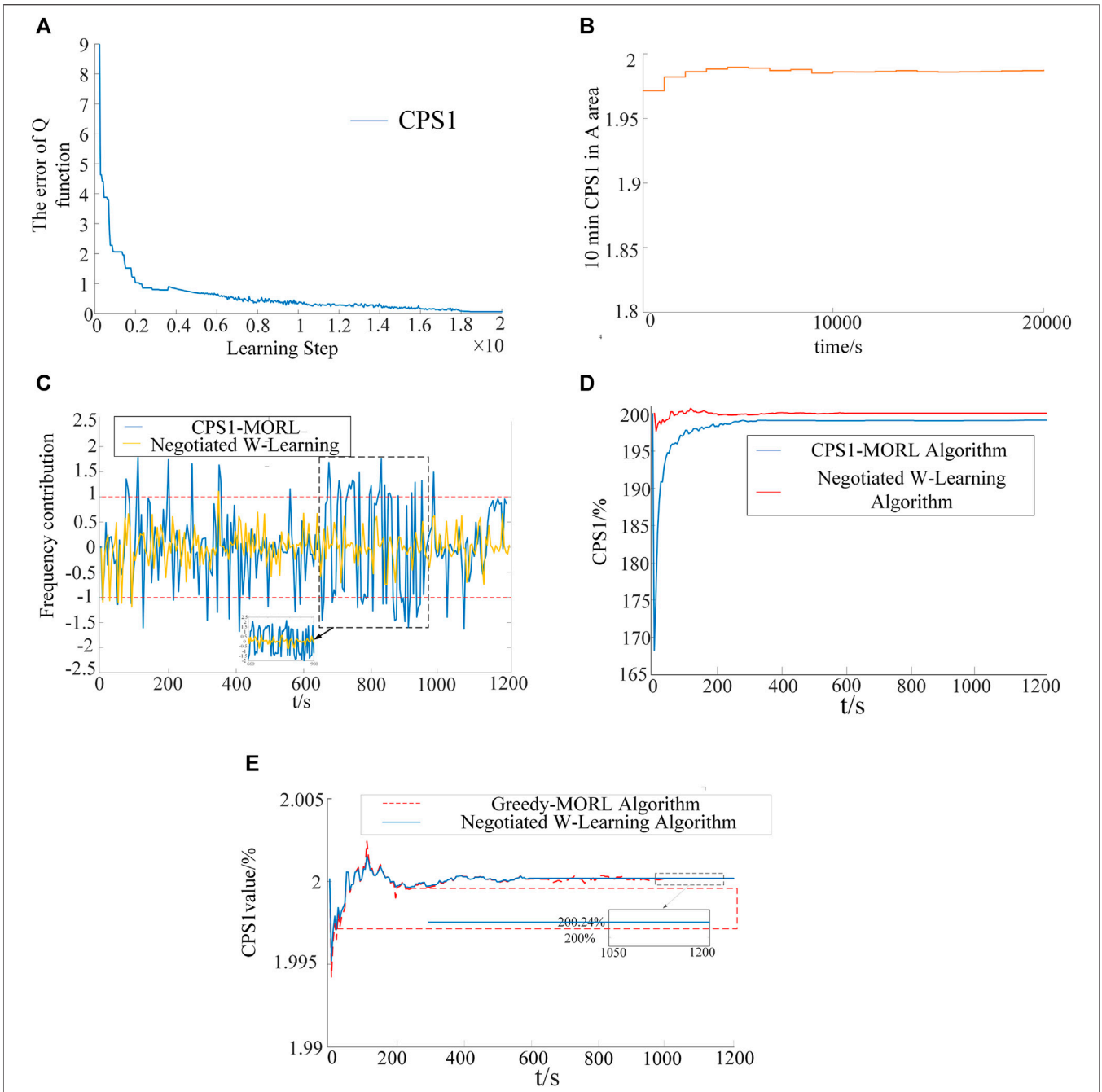

FIGURE 3 | Simulation results: **(A)** Differential convergence result of Q function under CPS1 objective, **(B)** $CPS1_{avg^{-}10-min}$ curve, **(C)** Self-contribution curve, **(D)** Curve reflecting the change of CPS1 value, **(E)** The curve of CPS1.

$$a_{cir} = \arg\max Q_{cir}(s, a) \tag{14}$$

Step 2: The remaining objective functions are calculated according to the following methods, as shown in 15:

$$W_i = \max Q_i(s, a) - Q_i(s, a_{cir}) \tag{15}$$

Step 3: Choose the maximum value of for other objective functions except the guide objective function, and compare it with W_{cir}. If $W_{i,\max} > W_{cir}$, the objective function which is corresponding to this maximum value of W_i should be selected as the new guidance objective function, the guidance value W_{cir} should be updated as the value of $W_{i,max}$, the corresponding action a should be made to be the new guidance action a_{cir}, and then go back to step 2 for repeated iterations until this condition is no longer met.

TABLE 1 | Simulation results under different algorithms.

Algorithms	Calculating time/s (pre-learning)	$\|\Delta f\|$/Hz	CPS1%	BAAL%
CPS1-MORL	12,031	0.0143	196	86.4
Coordinate Q-MORL	18,546	0.0132	197	96.2
Greedy-MORL	20,015	0.0129	199	97.2
Negotiated W-Learning	21,457	0.0064	200	98.5

If $W_i \le W_{cir}$ is obtained, record the guidance action a_{cir} and the guidance objective function at this time as the final value.

4 SIMULATION RESULTS

This paper builds a typical two-region interconnected power grid AGC model for controlling load frequency. The parameter settings of the two regions in the model system are the same, and the system base capacity is 1000 MW.

Figure 3A,B shows the pre-learning process of single CPS1 target and Negotiated W-Learning Algorithm. In the pre-learning stage, a continuous sinusoidal load disturbance with a period of 1,200 s, an amplitude of 100 MW and a duration of 20,000 s is applied to the A area, and a 2-norm Q function matrix $\|Q_t(s,a) - Q_{t-1}(s,a)\|^2 \le \zeta$ (ζ is a constant) is used as the standard for pre-learning to achieve the optimal strategy (Imthias Ahamed et al., 2002).

It can be seen from **Figure 3A** that after many iterations, the Q function tends to stabilize, reaching the optimal strategy for the CPS1 target. **Figure 3B** shows the average value of CPS1 ($CPS1_{\text{avg-10-min}}$) in area A every 10 min during the pre-learning process. It is found that the curve almost remains at a stable and acceptable value in the later stage, which shows that the Negotiated W-Learning algorithm has approached the optimal CPS1 control strategy. At the same time, the Q matrix corresponding to the target BAAL has also converged.

In addition, from the perspective of algorithm learning time, the four algorithms have been simulated for many times, and the average calculation time has been counted. See **Table 1** for details. Due to the difference in the number of optimization targets and the difficulty of calculating the coordination factor, the calculation time of the single target CPS1-MORL is the shortest. Since the CoordinateQ-MORL algorithm cannot fully explore the action set, its calculation time is the second. Compared with the global search algorithm Greedy-MORL, Negotiated W-Learning has gone through more search steps, so its time is the longest.

In order to further verify the adaptability of Negotiated W-Learning in the constantly changing power grid environment, this paper applies a random disturbance with a period of 1,200 s and an amplitude of 100 MW in area A. Four types of algorithms are set for comparison as follows.

Algorithm 1. Traditional single-objective reinforcement learning algorithm for intelligent frequency control based on CPS1 frequency control performance evaluation index (CPS1-MORL).

Algorithm 2. Multi-objective reinforcement learning algorithm for intelligent frequency control based on the traditional greedy strategy of multi-dimensional frequency control performance evaluation index and multi-objective Q function (Coordinate Q-MORL).

Algorithm 3. Under the traditional greedy strategy, this algorithm uses a cooperative reward function based on multi-dimensional frequency control performance evaluation indicators to achieve multi-objective reinforcement learning and intelligent frequency control algorithm (Greedy-MORL).

Algorithm 4. The Negotiated W-Learning algorithm proposed in this paper is based on the collaborative reward letter under the multi-dimensional frequency control performance evaluation index for multi-objective reinforcement learning and intelligent frequency control (Negotiated W-MORL).

4.1 Control Strategy Performance Analysis

Figure 3C shows the frequency deviation self-contribution degree ($\Delta f/\varepsilon$) and CPS1 index change curve of **Algorithm 1** and **Algorithm 4**. In this paper, the threshold is used for calculation, where $\mathcal{E}$ is 0.01. The frequency contribution degree has the ability to reflect the frequency quality of different algorithms. If the frequency contribution degree exceeds ± 1, it means that the frequency at this time has exceeded the prescribed limit 3ε. It can be seen that the frequency contribution curve of **Algorithm 1** exceeds the short-term index frequency continuous limit time specified in this article and has a steep drop in this interval, which will cause greater influence on system operation safety. However, the frequency contribution curve of **Algorithm 4** stays within the defined range. There are two main reasons for this phenomenon: One is that **Algorithm 4** controls the frequency by relaxing the weights of the two indicators in real time. If frequency fluctuations or "frequency drops" occur, the BAAL indicator will be given greater weight. If the frequency continuously exceeds the limit during the simulation period, CPS1 will be given a larger weight for regulation. The second is that **Algorithm 4** considers two indicators to participate in the evaluation of AGC control at the same time, while **Algorithm 1** only considers the impact of CPS1. At the same time, the CPS1 curve of **Algorithm 4** in **Figure 3D** fluctuates less throughout the simulation cycle, while the fluctuation of **Algorithm 1** is larger, which further proves that **Algorithm 4** is superior to **Algorithm 1** in terms of frequency control effect.

In summary, combining the BAAL and CPS1 indicators to constrain the system frequency can effectively improve the frequency quality of the system at the full time scale.

4.2 The Influence of Cooperative Reward Function on Frequency Control Performance

In order to verify the effectiveness of the collaborative reward function proposed in this paper, the control performance indicators of **Algorithm 2** and **Algorithm 3** can be compared. It can be seen that the control performance indicators of **Algorithm 3** are better than those of **Algorithm 2**. This is because the introduction of coordination factors between the multi-objective state-action value function may cause the agent to not fully explore the action set, leading to the omission of key actions, and the use of collaborative reward functions can effectively solve the above problems.

In summary, the introduction of a collaborative reward function can effectively improve the system frequency quality and various frequency performance indicators.

4.3 The Influence of Different Learning Strategies on Control Performance

In order to verify the effectiveness of **Algorithm 4** proposed in this paper, **Figure 3D** shows the CPS1 curve of **Algorithm 3** and **Algorithm 4**. It can be seen from **Figure 3E** that **Algorithm 4** has a faster convergence rate and a more stable fluctuation situation than **Algorithm 3** after the occurrence of load disturbance. This is because the Negotiated W-Learning strategy selects actions from global considerations, which effectively improves the traditional greedy strategy that is, easy to fall into the local optimal solution problem.

In summary, the global search strategy Negotiated W-Learning is more time-consuming than the local search strategies Greedy and CoordinateQ, but the search quality is higher.

5 CONCLUSION

This paper proposes a multi-intelligence frequency control strategy based on multi-dimensional evaluation criteria and cooperative reward function.

The simulation results show that: 1) Compared with the general algorithm, the Negotiated W-Learning algorithm can effectively improve the quality of the system frequency on the full time scale, and better explore the global action. 2) The collaborative reward function proposed in this paper can improve the linear weight of the traditional multi-objective Q function. In general, the intelligent AGC control strategy based on the collaboration of CPS1 and BAAL learning criteria proposed in this paper can effectively deal with the short-term power disturbance problem caused by the grid connection of new energy sources such as wind power, and improve the stability of the system.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

LZ put forward the main research points; LZ, YX completed manuscript writing and revision; JY completed simulation research; TX collected relevant background information; JC, ZL, TZ revised grammar and expression.

FUNDING

This manuscript was supported in part by the National Natural Science Foundation of China 52007103.

REFERENCES

Abouheaf, M., Gueaieb, W., and Sharaf, A. (2019). Load Frequency Regulation for Multi-Area Power System Using Integral Reinforcement Learning. *IET Generation, Transm. Distribution* 13 (19), 4311–4323. doi:10.1049/iet-gtd.2019.0218

Alhelou, H. H., Hamedani-Golshan, M.-E., Zamani, R., Heydarian-Forushani, E., and Siano, P. (2018). Challenges and Opportunities of Load Frequency Control in Conventional, Modern and Future Smart Power Systems: a Comprehensive Review. *Energies* 11 (10), 2497. doi:10.3390/en11102497

Arya, Y., and Kumar, N. (2017). Optimal Control Strategy–Based Agc of Electrical Power Systems: A Comparative Performance Analysis. *Optimal Control. Appl. Methods* 38 (6), 982–992. doi:10.1002/oca.2304

Imthias Ahamed, T. P., Rao, P. S. N., and Sastry, P. S. (2002). A Reinforcement Learning Approach to Automatic Generation Control. *Electric Power Syst. Res.* 63 (1), 9–26. doi:10.1016/s0378-7796(02)00088-3

Kumar, A., and Singh, O. (2019). "Recent Strategies for Automatic Generation Control of Multi-Area Interconnected Power Systems," in 2019 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE), Xi'an, China, April 25–27, 2019 (IEEE), 153–158. doi:10.1109/rdcape47089.2019.8979071

Li, Z., Jiang, W., Abu-Siada, A., Li, Z., Xu, Y., and Liu, S. (2020). Research on a Composite Voltage and Current Measurement Device for Hvdc Networks. *IEEE Trans. Ind. Electron.* 68, 8930-8941. doi:10.1109/tie.2020.3013772

Liu, H., Huang, K., Yang, Y., Wei, H., and Ma, S. (2018). Real-time Vehicle-To-Grid Control for Frequency Regulation with High Frequency Regulating Signal. *Prot. Control. Mod. Power Syst.* 3 (1), 1–8. doi:10.1186/s41601-018-0085-1

Liu, Y., Yang, N., Dong, B., Wu, L., Yan, J., Shen, X., et al. (2020). Multi-lateral Participants Decision-Making: A Distribution System Planning Approach with Incomplete Information Game. *IEEE Access* 8, 88933–88950. doi:10.1109/access.2020.2991181

Nathan, S., and Ballard, D. (2003). *Multiple-goal Reinforcement Learning with Modular Sarsa (0)*. Doctoral Dissertation Rochester: University of Rochester.

Shen, X., Ouyang, T., Yang, N., and Zhuang, J. (2021). Sample-based Neural Approximation Approach for Probabilistic Constrained Programs. *IEEE Trans. Neural Networks Learn. Syst.* 1-8. doi:10.1109/tnnls.2021.3102323

Shen, X., Ouyang, T., Khajorntraidet, C., Li, Y., Li, S., and Zhuang, J. (2021). Mixture Density Networks-Based Knock Simulator. *IEEE/ASME Trans. Mechatronics*, 1. doi:10.1109/tmech.2021.3059775

Shen, X., and Raksincharoensak, P. (2021). Pedestrian-aware Statistical Risk Assessment. *IEEE Trans. Intell. Transportation Syst.*, 1–9. doi:10.1109/tits.2021.3074522

Shen, X., and Raksincharoensak, P. (2021). Statistical Models of Near-Accident Event and Pedestrian Behavior at Non-signalized Intersections. *J. Appl. Stat.*, 1–21. doi:10.1080/02664763.2021.1962263

Shen, X., Zhang, Y., Sata, K., and Shen, T. (2020). Gaussian Mixture Model Clustering-Based Knock Threshold Learning in Automotive Engines. *IEEE/ASME Trans. Mechatronics* 25 (6), 2981–2991. doi:10.1109/tmech.2020.3000732

Shen, X., Zhang, X., Ouyang, T., Li, Y., and Raksincharoensak, P. (2020). Cooperative Comfortable-Driving at Signalized Intersections for Connected and Automated Vehicles. *IEEE Robotics Automation Lett.* 5 (4), 6247–6254. doi:10.1109/lra.2020.3014010

Shen, X., Zhang, Y., Shen, T., and Khajorntraidet, C. (2017). Spark advance self-optimization with knock probability threshold for lean-burn operation mode of si engine. *Energy* 122, 1–10. doi:10.1016/j.energy.2017.01.065

Wang, C., and James, D. M. C. (2013). Impact of Wind Power on Control Performance Standards. *Int. J. Electr. Power Energ. Syst.* 47, 225–234. doi:10.1016/j.ijepes.2012.11.010

Wang, H., Lei, Z., Zhang, X., Peng, J., and Jiang, H. (2019). Multiobjective Reinforcement Learning-Based Intelligent Approach for Optimization of Activation Rules in Automatic Generation Control. *IEEE Access* 7, 17480–17492. doi:10.1109/access.2019.2894756

Watkins, C. J. C. H., and Dayan, P. (1992). Q-learning. *Machine Learn.* 8 (3-4), 279–292. doi:10.1023/a:1022676722315

Xi, L., Lu, Y., Xu, Y., Wang, S., and Chen, X. (2019). A Novel Multi-Agent Ddqn-Ad Method-Based Distributed Strategy for Automatic Generation Control of Integrated Energy Systems. *IEEE Trans. Sustain. Energ.* 11 (4), 2417–2426. doi:10.1109/tste.2019.2958361

Xi, L., Zhou, L., Xu, Y., and Chen, X. (2020). A Multi-step Unified Reinforcement Learning Method for Automatic Generation Control in Multi-Area Interconnected Power Grid. *IEEE Trans. Sustain. Energ.* 12 (2), 1406–1415. doi:10.1109/tste.2020.3047137

Xie, Y., Zhang, H., Li, C., and Sun, H. (2017). Development Approach of a Programmable and Open Software Package for Power System Frequency Response Calculation. *Prot. Control. Mod. Power Syst.* 2 (1), 1–10. doi:10.1186/s41601-017-0045-1

Yang, N., Huang, Y., Hou, D., Liu, S., Ye, D., Dong, B., et al. (2019). Adaptive Nonparametric Kernel Density Estimation Approach for Joint Probability Density Function Modeling of Multiple Wind Farms. *Energies* 12 (7), 1356. doi:10.3390/en12071356

Yang, N., Liu, S., Deng, Y., and Xing, C. (2021). An Improved Robust Scuc Approach Considering Multiple Uncertainty and Correlation. *IEEJ Trans. Electr. Electron. Eng.* 16 (1), 21–34. doi:10.1002/tee.23265

Yang, N., Yang, C., Wu, L., Shen, X., Jia, J., Li, Z., et al. (2021). Intelligent Data-Driven Decision-Making Method for Dynamic Multi-Sequence: An E-Seq2seq Based Scuc Expert System. *IEEE Trans. Ind. Inform.* 1. doi:10.1109/tii.2021.3107406

Yang, N., Ye, D., Zhou, Z., Cui, J., Chen, D., and Wang, X. (2018). Research on Modelling and Solution of Stochastic Scuc under Ac Power Flow Constraints. *IET Generation, Transm. Distribution* 12 (15), 3618–3625.

Yu, T., Wang, Y. M., Ye, W. J., Zhou, B., and Chan, K. W. (2011). Stochastic Optimal Generation Command Dispatch Based on Improved Hierarchical Reinforcement Learning Approach. *IET generation, Transm. distribution* 5 (8), 789–797. doi:10.1049/iet-gtd.2010.0600

Zhang, Lei., and Luo, Yi. (2018). Combined Heat and Power Scheduling: Utilizing Building-Level thermal Inertia for Short-Term thermal Energy Storage in District Heat System. *IEEJ Trans. Electr. Electron. Eng.* 13 (6), 804–814. doi:10.1002/tee.22633

Zhu, B., Ding, F., and Don, M. V. (2019). Coat Circuits for Dc–Dc Converters to Improve Voltage Conversion Ratio. *IEEE Trans. Power Electron.* 35 (4), 3679–3687.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.