Parallel System Based Quantitative Assessment and Self-evolution for Artificial Intelligence of Active Power Corrective Control

Tianyun Zhang, *Student Member*, *IEEE*, Jun Zhang, *Senior Member*, *IEEE*, Feiyue Wang, *Fellow*, *IEEE*, Peidong Xu, *Student Member*, *IEEE*, Tianlu Gao, *Student Member*, *IEEE*, Haoran Zhang, and Ruiqi Si

***Abstract*—In artificial intelligence (AI) based-complex power system management and control technology, one of the urgent tasks is to evaluate AI intelligence and invent a way of autonomous intelligence evolution. However, there is, currently, nearly no standard technical framework for objective and quantitative intelligence evaluation. In this article, based on a parallel system framework, a method is established to objectively and quantitatively assess the intelligence level of an AI agent for active power corrective control of modern power systems, by resorting to human intelligence evaluation theories. On this basis, this article puts forward an AI self-evolution method based on intelligence assessment through embedding a quantitative intelligence assessment method into automated reinforcement learning (AutoRL) systems. A parallel system based quantitative assessment and self-evolution (PLASE) system for power grid corrective control AI is thereby constructed, taking Bayesian Optimization as the measure of AI evolution to fulfill autonomous evolution of AI under guidance of their intelligence assessment results. Experiment results exemplified in the power grid corrective control AI agent show the PLASE system can reliably and quantitatively assess the intelligence level of the power grid corrective control agent, and it could promote evolution of the power grid corrective control agent under guidance of intelligence assessment results, effectively, as well as intuitively improving its intelligence level through self-evolution.**

***Index Terms*—AI quantitative intelligence assessment and self-evolution, automated reinforcement learning, Bayesian optimization, corrective control, parallel system.**

NOMENCLATURE

k	Iteration times during evolution.
K	Total times of iteration.
$p^{(k)}$	Hyperparameters of the power grid corrective control AI agents for kth iteration.
$T^{(k)}$	Power grid corrective control AI agent obtained in kth iteration.
$I^{(k)}, I_p$	Comprehensive intelligence level of $T^{(k)}$.
T^*	The best power grid corrective control AI agent.
p^*	Hyperparameters of the best power grid corrective control AI agent.
I^*	The highest comprehensive intelligence level.
J	Reward function.
n_m	Testing scenarios for intelligence assessment.
m	Intelligence indexes (i-indexes) selected for intelligence assessment.
A	Artificial power system.
R	Real power system.
t	Time.
$F_t^{A,(k)}$	State transition functions of the artificial power system at t.
$H_t^{A,(k)}$	Observation functions of the artificial system.
$U_t^{R,(k)}$	Actions generated by $T^{(k)}$ in the real power system at t.
$U_t^{A,(k)}$	Actions generated by $T^{(k)}$ in the artificial power system at t.
$Y_t^{R,(k)}$	Observations of $T^{(k)}$ in the real power system at t.
$Y_t^{A,(k)}$	Observations of $T^{(k)}$ in the artificial power system at t.
$D(Y_{1:t}^{R,(k)}, Y_{1:t}^{A,(k)})$	Distance between the real and the artificial system observations.
P	Active power.
Q	Reactive power.
B	Connected bus.
Pre	Generator re-dispatching.
ρ	Load ratio.
$X_t^{A,(k)}$	Artificial power system states at t.
$X_{t,P}^{A,(k)}$	Active power states of the power grid at t.
$X_{t,Q}^{A,(k)}$	Reactive power states of the power grid at t.
$X_{t,B}^{A,(k)}$	Connected bus of power devices in substations at t.
$X_{t,\mathrm{Pre}}^{A,(k)}$	Re-dispatching amount of each generator at t.
$X_{t,\rho}^{A,(k)}$	Load ratio of power lines at t.

Manuscript received January 13, 2023; revised June 1, 2023; accepted August 3, 2023. Date of online publication December 28, 2023; date of current version December 25, 2023. This work was supported by the National Key R&D Program of China [grant number 2018AAA0101504]. The work is funded by the Science and Technology Project of SGCC (State Grid Corporation of China): fundamental theory of human-in-the-loop hybrid-augmented intelligence for power grid dispatch and control.

T. Y. Zhang, J. Zhang (corresponding author, email: jun.zhang.ee@whu.edu.cn; ORCID: https://orcid.org/0000-0001-6908-2671), P. D. Xu, T. L. Gao, H. R. Zhang and R. Q. Si are with the School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, Hubei, China.

F. Y. Wang is with the State Key Laboratory for Management and Control of Complex System, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.

DOI: 10.17775/CSEEJPES.2023.00190

$Y_{t,P}^{A,(k)}$	Active power observations of the power grid at t.
$Y_{t,Q}^{A,(k)}$	Reactive power observations of the power grid at t.
$Y_{t,B}^{A,(k)}$	Connected bus observations of power devices in substations at t.
$Y_{t,\mathrm{Pre}}^{A,(k)}$	Re-dispatching observations of each generator at t.
$Y_{t,\rho}^{A,(k)}$	Load ratio observations of power lines at t.
N_{line}	Number of power lines in the power grid.
$E_m(Y_1^{A,(k)}, T^{(k)}, n_m)$	Index intelligence score.
w_m	Index weight.
M	Total number of the selected intelligence indexes.
t_{end}	Continuous operation time before the blackout.
t_{E}	Predetermined operation time of the agent in the scenarios.
$C_{\mathrm{op}}(n_m)$	Operational cost.
$c_{\mathrm{loss}}(Y_{t,P,\mathrm{line}}^{A,(k)}, t)$	Energy loss cost.
$c_{\mathrm{Pre}}(Y_{t,\mathrm{Pre}}^{A,(k)}, t)$	Cost of the generation re-dispatching.
$c_{\mathrm{b}}(Y_{t,P}^{A,(k)}, t)$	Blackout cost.
N	Total number of the index scoring equations.
n	Serial number of the index scoring equations.
s	Current training episode.
S	Total training episodes.
$E_{m,n}$	Index scores of the nth scoring equation in n_m of the i-index m.
$E''_{m,n}$	Normalized index scores through Z-score normalization.
$E'_{m,n}$	Normalized index scores through Logistics normalization.
$r_{m,n}$	Proportion of $E'_{m,n}$.
e_n	Information entropy of the nth scoring function.
w_n	Weight of the nth scoring function.
His	Search history.
$\alpha(p)$	Acquisition function.
$\mu(I_p)$	Posterior mean of I_p.
$\sigma(I_p)$	Standard deviation of I_p.
$\Phi(\cdot)$	Cumulative distribution function of the standard normal distribution.
$\varphi(\cdot)$	Probability density function of the standard normal distribution.
line	Transmission lines.
$C(\cdot)$	Adjustment cost function.
G	Generator power.
L	Load power.
$Y_{t,P,\mathrm{line}}^{A,(k)}$	Active power of the transmission lines.
$\Delta X_{P,G}^{A,(k)}$	Amount of generator re-dispatching in the power system.
$\Delta X_{P,L}^{A,(k)}$	Amount of load-shedding in the power system.
$X_{t,P,G}^{A,(k)}$	Generator power at t.
$X_{t,P,L}^{A,(k)}$	Load power at t.
$P_{\max}$	The maximum value of the line active power.
$N_{U,\mathrm{line}}$	Number of actions of the line switching.
$N_{U,\mathrm{bus}}$	Number of actions of the bus switching.
$N_{U,C}$	Number of actions of the allowable topology adjustment.
$P_{G\max}$	Upper limits of generator outputs.
Ra_{up}	Upper limits of generator ramping capabilities.
$P_{G\min}$	Lower limits of generator outputs.
Ra_{down}	Lower limits of generator tamping capabilities.

I. INTRODUCTION

ARTIFICIAL Intelligence (AI) agents for power system management and control are generally complicated to evaluate, needless to say, and compare them with unified standards [1], [2]. Meanwhile, this evaluation and comparison would help dispatchers find proper actions for power system management and control, improve stability and reliability of power system operation, and establish an AI agent with efficient power system management and control abilities. Besides, to control the power system operation status more effectively, while reducing workload of dispatchers, self-evolution is introduced in power system management and control AI, which could be fulfilled by a reinforcement learning algorithm. Reinforcement learning (RL) faces two problems when dealing with AI construction and evolution. The first aporia is how to quickly and efficiently explore AI models with better performance in real scenes, where it is very expensive to manually find hyperparameters. Another thorny problem is the flexibility of RL algorithms requires improvement in a field in need of massive calculations. Currently, these problems can be solved by Automated Reinforcement Learning (AutoRL) [3]. However, when attempting to understand and interpret whether AutoRL evolutionary result is the best agent, AutoRL would encounter the same problems as AI agents.

Among research on applications of AI in power system management and control, a considerable number of them focus on power grid corrective control AI agents. These agents perform active power corrective control of power systems, to ensure safe and stable operation of power grids through appropriate actions by eliminating or reducing overload on power lines. These actions include generator re-dispatching, load-shedding, and network topology adjustment. At present, research on power grid corrective control mainly focus on adjusting power system states when the system encounters abnormal operation states, by taking traditional optimization methods or AI. These studies aim at improving transient stability of modern power grids through different corrective control actions [4], balancing preventive and corrective control of the transmission network [5], and dealing with uncertainty of power grids through corrective control [6]. Rapid development of AI technology finds utilizing agents to solve the power grid corrective control problem has become an effective method, among which applications of RL in related problems account for the majority [7]–[11]. Power grid corrective control agent constructed in this research could reduce powerline overload with relatively minimal costs. However, research on power grid corrective control AI agent still stays in the area of effective design and performance improvement, while little attention has been drawn to agent intelligence assessment and self-evolution. A quantitative intelligence assessment could help dispatchers find or establish an agent with efficient active power corrective control actions and ability, while an intelligence assessment-based AI self-evolution could improve accuracy of the agent in power grid control, thus preventing accidents caused by failures of power grid corrective actions. Thereby, this article intends to construct an intelligence assessment and self-evolution system for power grid corrective control AI.

Intelligence assessment refers to a method of describing the intelligence level of AI agents by virtue of subjective or objective measurements. At present, research on AI and its intelligence evaluation are mainly subjective and qualitative. For example, methods such as the Turing test [12], similarity between machine behaviors and human behaviors [13], and expert scoring [14] are taken to evaluate intelligence level of AI. These methods rely highly on expert experiences, thereby, being subjective ways to reflect intelligence level of AI. Quantitative assessment of AI intelligence level has been a difficult-to-crack task in the field of AI. So far, only two studies cover the area. One proposes an evaluation standard for objective assessment of machine intelligence based on Abstraction and Reasoning Corpus (ARC) dataset, a benchmark for general artificial intelligence [1], while the other offers the concept and method of machine intelligence evaluation based on video games [15]. These standards and methods, however, are still at the theoretical level, only applicable in areas of graphics or video games. A general objective and quantitative evaluation framework of machine intelligence are barely available [16], let alone intelligence of power system management and control agent. By this token, urgently, we need to propose a unified and easy-to-understand method to better explain relations among agents' observations and actions, and further compare their intelligence levels.

In addition to AI intelligence assessment, AI self-evolution attracts great attention. Automated Reinforcement Learning (AutoRL) is one of the classic ways for AI self-evolution. AutoRL is a specific implementation of Automated Machine Learning (AutoML) specifically for evolution of RL agent, to fulfill self-evolution of agents by independently adjusting network architecture, hyperparameters, and algorithms of RL agents [3]. AutoML is a data-driven AI generation and optimization system that automates the whole machine learning process, aiming to greatly reduce development costs of AI agents with better performance [17]–[23]. The widely accepted optimization methods of AutoML involve grid search, random search, evolutionary algorithm [20], Bayesian Optimization [17]–[19], [23], etc. As a popular self-evolution method, Bayesian Optimization (BO) is applied in industry and scientific experiments in light of its concept of sequential decision-making. Specifically, Bayesian Optimization has been used to adjust hyperparameters of AlphaGo to improve its winning rate [3], [24]; and an information-theoretic framework for constrained BO was constructed to solve global black-box optimization problems [25]. Other researches include applying BO to optimize the hyperparameter set [26], and conducting BO in conditional hyperparameter spaces [27], to solve the problem of algorithm selection and hyperparameter optimization. Reference [28] proposes Ensemble Bayesian Optimization, which covers problems of large-scale observations, high-dimensional input spaces, and batch query selections, with balanced quality and diversity. This research barely solves the problem of AI evaluation and comparison, thus a unified AI assessment method is necessary to be introduced into AutoRL and BO.

Taking all these considerations, to achieve self-evolution based on quantitative intelligence assessment for power system control and management agent, this article intends to build a Parallel System based Quantitative Assessment and Self-Evolution (PLASE) system for the power grid corrective control AI agent. The PLASE system includes a quantitative intelligence assessment index system by imitating human intelligence evaluation methods and theories, to quantitatively assess AI intelligence. On this basis, the PLASE system offers an intelligence assessment-based AI self-evolution method, with combined efforts of BO and AutoRL, mainly targeting deep reinforcement learning (DRL) agent for the power grid corrective control, and agent self-evolution. Finally, this article verifies the proposed PLASE system by experiments on a 36-bus power system. Results show the intelligence level of different agents can be assessed by the quantitative intelligence assessment index system. The intelligence level of power grid corrective control agents can be improved by adjusting their hyperparameters.

This article is organized as follows. Section II introduces the objectives of the power grid corrective control problem, and corresponding AI agent. Section III explores the PLASE system. Based on the parallel system framework, Section III briefly describes construction of the PLASE system following the ACP approach, as well as its operating principle. Section IV is devoted to experiments to verify the PLASE system can quantitatively assess the intelligence level of the power grid corrective control AI, and achieve AI self-evolution. Section V concludes this article.

II. The Power Grid Corrective Control Problem

The main goal of the power grid corrective control problem is to minimize power grid adjustment cost and ensure maximum line transmission power while keeping the system working. In this process, disturbance to the power system caused by these adjustments should remain as low as possible. The objective function and constraints of the power grid corrective control problem can be expressed in (1) and (2), assuming there is no cost for network topology adjustment.

$$\max Y_{t,P,\text{line}}^{A,(k)} + \min C\left(\left|\Delta X_{P,G}^{A,(k)}\right|, \Delta X_{P,L}^{A,(k)}\right) \tag{1}$$

$$\begin{aligned} \textit{s.t.}\ & Y_{t,P,\text{line}}^{A,(k)} \le P_{\max} \\ & N_{U,\text{line}} + N_{U,\text{bus}} \le N_{U,C} \\ & \left|\Delta X_{P,G}^{A,(k)}\right| \le \min\left(P_{G\max} - X_{t,P,G}^{A,(k)}, Ra_{\text{up}}\right) \\ & \left|\Delta X_{P,G}^{A,(k)}\right| \le \min\left(X_{t,P,G}^{A,(k)} - P_{G\min}, Ra_{\text{down}}\right) \end{aligned} \tag{2}$$

where $Y_{t,P,\text{line}}^{A,(k)}$ is active power of the transmission lines; $\Delta X_{P,G}^{A,(k)}$ is amount of generator re-dispatching in the power system, $X_{t,P,G}^{A,(k)}$ is generator power at time t, $\Delta X_{P,G}^{A,(k)} =$

$X_{t,P,G}^{A,(k)} - X_{t-1,P,G}^{A,(k)}$; $\Delta X_{P,L}^{A,(k)}$ is amount of load-shedding in the power system, $X_{t,P,L}^{A,(k)}$ is load power at t, $\Delta X_{P,L}^{A,(k)} = X_{t,P,L}^{A,(k)} - X_{t-1,P,L}^{A,(k)}$; $P_{\max}$ is maximum value of line active power; $N_{U,\text{line}}$ is number of actions of line switching, $N_{U,\text{bus}}$ is number of actions of bus switching, $N_{U,C}$ is number of actions of allowable topology adjustment; $P_{G\max}$ is upper limits of generator outputs, Ra_{up} is upper limits of generator ramping capabilities; $P_{G\min}$ is upper limits of generator outputs, Ra_{down} is upper limits of generator ramping capabilities.

Power grid corrective control agent $T^{(k)}$ is thereby established on the basis of the Double Dueling graph Deep Q Network (DDDQN) grounded by DRL. Architecture of power grid corrective control agent is shown in Fig. 1.

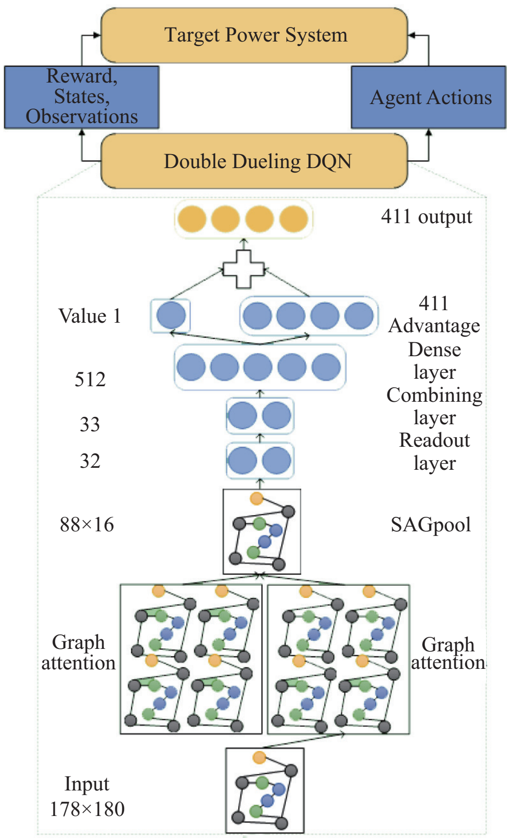

Fig. 1. Agent architecture.

Power grid corrective control agent is designed to solve the corrective control problem shown in (1) and (2) while preserving long-time operation of the power system. The agent should gain rewards according to previous and current power system operation states, to select more efficient actions while maintaining power grid stability. Therefore, immediate reward of the power grid corrective control agent is designed, as shown in (3). If power system fails due to improper control or action of agent at time t, reward would be minimum. When there is no data for power flow calculation at t, there will be $J = 0$. Under other circumstances, the reward will be calculated by performance evaluation function for $T^{(k)}$ at t, as shown in (4). This function aims at deciding validity of corrective control actions given by $T^{(k)}$, by evaluating available transfer capabilities (ATC) of target power system, to determine agent corrective control effect for target power system. In (4), certain penalties, gained by experimentation, are given to loaded power lines and overloaded power lines, which would have great impacts on power transmission [8].

$$J = \begin{cases} 0, & \text{if no more data} \\ -2, & \text{if the system fails} \\ \sum_{i=1}^{N_{\text{Line}}} J^*/100, & \text{otherwise} \end{cases} \tag{3}$$

$$\begin{aligned} J^* = {} & \max\left(0, (1-(Y_{t,\rho,i}^{A,(k)})^2)\right) - 10\max\left(0, Y_{t,\rho,i}^{A,(k)} - 1\right) \\ & - 5\max\left(0, Y_{t,\rho,i}^{A,(k)} - 0.9\right) \end{aligned} \tag{4}$$

III. THE PLASE SYSTEM

The Parallel System based Quantitative Assessment and Self-Evolution (PLASE) system for power grid corrective control AI resorts to parallel systems to realize power grid corrective control AI self-evolution based on intelligence assessment results. This section contributes to construction of the PLASE system from four perspectives: basic structure of the PLASE system, reinforcement learning agent for power grid corrective control, intelligence assessment, and self-evolution.

A. The Brief Discussion and the Objective Functions of the PLASE System

Current power system control agents are trained and evolved in artificial systems such as power system analysis and simulation software (like PSASP and PSEE). Therefore, to systematically explain intelligence assessment and self-evolution of power grid corrective control agents in the PLASE system, it is necessary to introduce the parallel system principle.

The parallel system principle is a recently developed complex system modeling and analysis paradigm, to accomplish complex system management and control through virtual-real interactions and parallel executions. A parallel system is a combination of the real complex system, artificial power system, and related auxiliary to undertake tasks of training, evaluation, and control. Through training, evaluation, management, and control of artificial systems, the parallel system obtains an evolution scheme of the real system to achieve the optimization goal and evolution of the real system. In the operation process, the parallel system follows the ACP method [29], in which **A** represents the artificial system generated in the process of constructing a parallel system; **C** refers to the computational experiments, which would compare, analyze, predict, and evaluate the real system and the artificial system; **P** is parallel execution, namely, simultaneously operating real and artificial system, realizing management and control of the real complex system through virtual and real interaction, and identifying optimal control schemes for settling complex system problems. ACP method implements three measures, being respectively "Learning and Training", "Experiment and Evaluation", and "Management and Control".

Based on the ACP method, the PLASE system would follow the process below to realize intelligence assessment and self-evolution of power grid corrective control AI agent. First of all, the PLASE system establishes the artificial power system of target power grid based on actual power system, while constructing and training AI agent to perform active power corrective control task of the power system through “Learning and Training”. Then, the intelligence level of the trained agent is quantitatively assessed by “Experiment and Evaluation”. Finally, through “Management and Control”, an evolved power grid corrective control agent is generated based on existing agents and their intelligence level. The basic structure of the PLASE system is shown in Fig. 2.

In the PLASE system, evolution of the power grid corrective control agent is mainly fulfilled by AutoRL and BO. According to the goal of AutoRL [3] and the basic principle of Bayesian Optimization [30], the problem of the power grid corrective control AI self-evolution based on intelligence assessment, studied in this article, is the process of finding power grid corrective control AI agent $T^{(k)}$ and its hyperparameters $p^{(k)}$ in the search space, which optimizes comprehensive intelligence level $I^{(k)}$ in finite iterations under the same intelligence indexes (i-indexes) and testing scenarios. In AI evolution, targets and constraints of power grid corrective control problems solved by AI should be satisfied. In short, the objective of the PLASE system is to find T^{*} and corresponding p^{*} satisfying following equations:

$$T^{*} = \underset{1\leq k\leq K}{\arg\max}\, I^{(k)}\big(T^{(k)}(p^{*}), n_m\big) \tag{5}$$

$$p^{*} = \underset{1\leq k\leq K}{\arg\max}\, J\big(T^{(k)}(p^{(k)})\big) \tag{6}$$

B. Reinforcement Learning Agent

The general solution process for the artificial system in the parallel system is shown in (7).

$$\begin{aligned} &\big\{\hat{\mathrm{F}}_t^{A,(k)}, \hat{\mathrm{H}}_t^{A,(k)}, \hat{\mathrm{U}}_t^{A,(k)}, \hat{\mathrm{U}}_t^{R,(k)}\big\} = \\ &\qquad \min_{\mathrm{F}_t^{A,(k)}, \mathrm{H}_t^{A,(k)}, \mathrm{U}_t^{A,(k)}, \mathrm{U}_t^{R,(k)}} D\big(Y_{1:t}^{R,(k)}, Y_{1:t}^{A,(k)}\big) \\ &\textit{s.t.}\ \text{Constraints}\big(\mathrm{F}_t^{A,(k)}, \mathrm{H}_t^{A,(k)}, \mathrm{U}_t^{A,(k)}, \mathrm{U}_t^{R,(k)}\big) \end{aligned} \tag{7}$$

For real and artificial power systems under investigation in this article, state transition function of artificial power system $F_t^{A,(k)}$ and observation function of artificial power system $H_t^{A,(k)}$ are known, actions in the artificial power system $U_t^{A,(k)}$ and real power system $U_t^{R,(k)}$ need to be provided by the AI agent $T^{(k)}$ configured in the target power system, and distance between the real and the artificial observations $D(Y_{1:t}^{R,(k)}, Y_{1:t}^{A,(k)})$ could be ignored. Typically, since training and evolution of the power grid corrective control agent are accomplished in the artificial power system in the parallel system, only $U_t^{A,(k)}$ needs to be considered. When the agent is utilized in the real power system, $U_t^{R,(k)}$ needs to be considered.

Since $U_t^{A,(k)}$ and $U_t^{R,(k)}$ are provided by the power grid corrective control agent $T^{(k)}$ in each iteration, $T^{(k)}$ generated by the PLASE system could be expressed by (8).

$$T^{(k)} = J\big(X_t^{A,(k)}, Y_t^{A,(k)}, Y_t^{R,(k)}, p^{(k)}\big) \tag{8}$$

The process of using reward function J to control and train $T^{(k)}$, achieved by “Learning and Training” in the PLASE system, could be expressed by (9).

$$\{\hat{T}^{(k)}\} = \max_{T^{(k)}}\{J\} \tag{9}$$

Specifically, the basic structure of $T^{(k)}$ is shown in Fig. 1, J is represented by (3), and the power grid corrective control problem solved in this article is expressed in Section II.

C. The Intelligence Assessment

To achieve objective and quantitative intelligence assessment of the power grid corrective control AI agent, this article proposes a quantitative AI intelligence assessment index system by imitating a human intelligence testing method, whose basic intelligence assessing steps are shown in Fig. 3.

Step 1: Constructing quantitative AI intelligence assessment index system, and selecting i-indexes m according to required tasks and capabilities of the target power grid corrective control AI agent.

AI agents are constructed by imitating human intelligence, with the promise they shall have intelligence similar to human beings in some aspects. Therefore, it is feasible to imitate the quantitative evaluation method of human intelligence to realize quantitative intelligence assessment of AI. The research on human intelligence evaluation has been relatively well-developed, and various human intelligence test methods have been established, such as the test method based on Gardner’s multiple intelligence theory [31] and the Raven intelligence

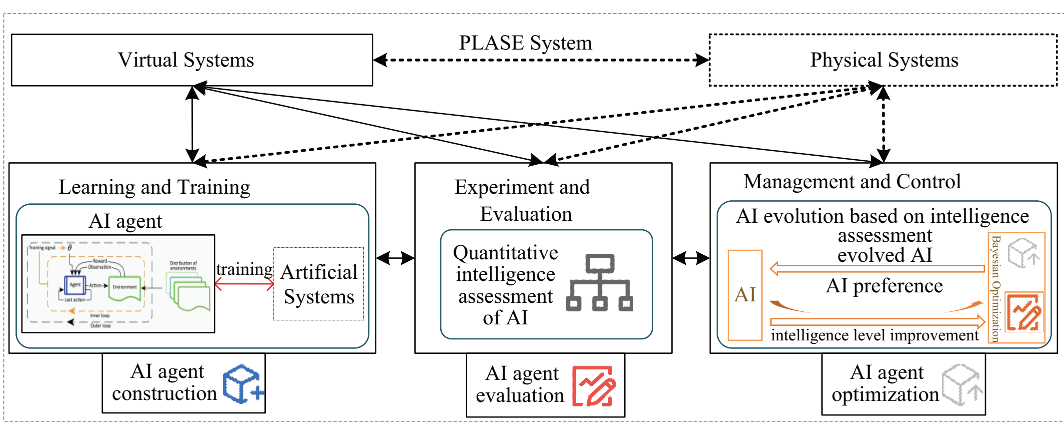

Fig. 2. The parallel system and the ACP method.

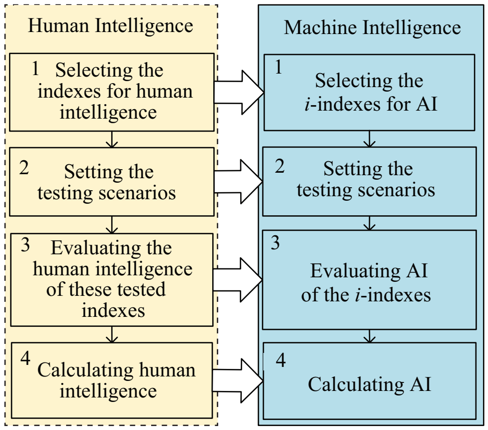

Fig. 3. The basic process of quantitative intelligence assessment.

test [32]. These methods can provide references for AI intelligence assessment. It is known that human intelligence is divided into corresponding intelligence indexes according to human abilities related to intelligence. These human intelligence tests often evaluate human intelligence indexes through corresponding examinations and tests. Test scores of various intelligence indexes are used to objectively and quantitatively evaluate human intelligence. On these grounds, quantitative assessment of machine intelligence can also be realized through a similar index-based methodology.

Machine intelligence refers to abilities of AI agents to achieve their design goals in different scenarios. According to the intelligence index division of Gardner's multiple intelligence theory and human cognitive division theory [33], machine intelligence could also be divided into multiple intelligence indexes through AI capabilities and tasks. By analyzing capabilities an AI agent should have, 35 intelligence indexes are proposed for machine intelligence. These 35 indexes are further classified and summarized in the following three aspects: similarity and superiority compared with human beings, completion of tasks, and intelligence level possessed by agents to meet their design goals. In this process, the 35 indexes are divided into refined indexes and comprehensive indexes, considering some indexes share similar intelligent features. Refined indexes are a collection of intelligence indexes with the same intelligent features, while comprehensive indexes do not contain these intelligent features. The quantitative intelligence assessment index system is constructed accordingly, as shown in Fig. 4. In Fig. 4, the proposed 35 intelligence indexes are marked in red, containing 26 refined indexes and 9 comprehensive indexes. When assessing AI intelligence, a group of appropriate i-indexes will be selected from this index system.

Step 2: Generating testing scenarios n_m according to i-indexes m. When setting testing scenarios, specific index scenario requirements, similarity between training and testing scenarios, and difficulty of the testing scenarios will be decided through abilities represented by selected i-indexes. Testing scenarios will then be generated by these requirements.

Step 3: Obtaining operation states observations $Y_t^{A,(k)}$ of the power grid corrective control AI agent in testing scenarios through computational experiment, evaluating agent intelligence scores of m, $E_m(Y_t^{A,(k)}, T^{(k)}, n_m)$, through n_m and index scores of selected index scoring equations $E_{m,n}$. Calculation of $E_{m,n}$ is fulfilled by selected scoring equations and $Y_t^{A,(k)}$. The specific index scoring equation used in this article is discussed in Case Study. If only one kind of scoring equation is chosen, mean value of the scores of the index scoring equation $E_{m,n}$ equals to $E_m(Y_t^{A,(k)}, T^{(k)}, n_m)$. If multiple kinds of scoring equations are chosen, the detailed calculation process, using the entropy method, is shown in Algorithm 1.

Step 4: Computing comprehensive intelligence level $I^{(k)}$ of $T^{(k)}$ according to agent intelligence scores of i-indexes m gained in Step 3. The formula for computing intelligence level is shown in (10). If only one i-index is chosen in Step 1, its weight would default to 1. If multiple i-indexes are chosen in Step 1, weights of i-indexes would be gained by the entropy

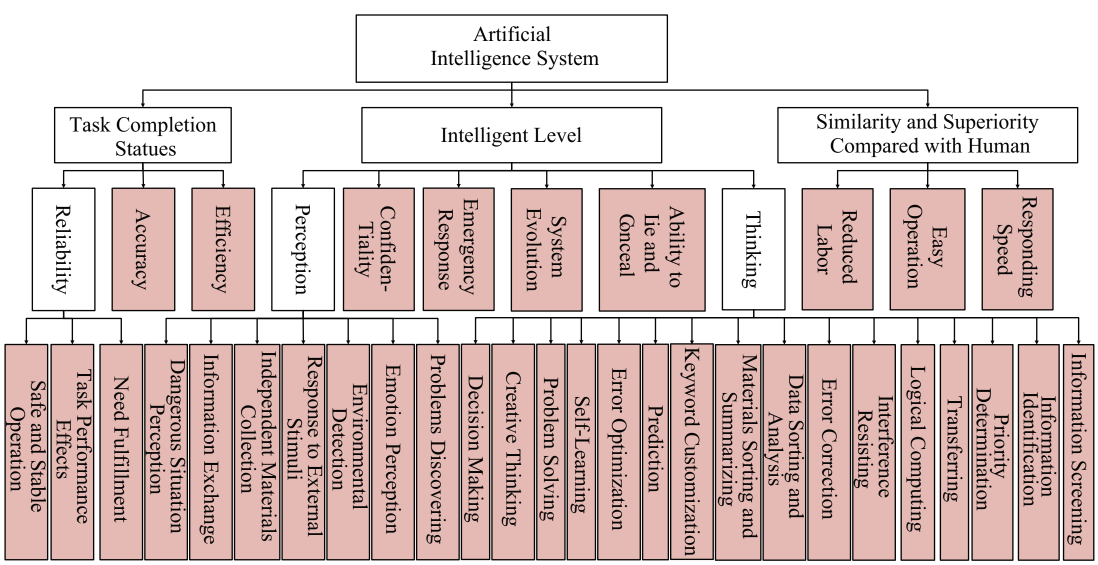

Fig. 4. Quantitative intelligence assessment index system.

Algorithm 1: Algorithm for the Intelligence scores of i-indexes using the entropy method

1 Calculating index scores in each scenario $E_{m,n}$ using scoring functions
2 Normalizing $E_{m,n}$ as $E'_{m,n}$ using Z-score $E''_{m,n} = (E_{m,n} - \mu(E_{m,n}))/\sigma(E_{m,n})$ and Logistic $E'_{m,n} = \frac{1}{1+e^{-E''_{m,n}}}$
3 Getting proportion r_{mn} for normalized index scores $E'_{m,n}$ by $r_{mn} = E'_{m,n}/\sum_{m=1}^{M} E'_{m,n}$
4 Calculating information entropy $e_n = \frac{-1}{\ln(M)} \sum_{m=1}^{M} r_{mn} \ln(r_{mn})$
5 Getting weights $w_n = (1-e_n)/\sum_{n=1}^{N}(1-e_n)$
6 Getting total index scores $E_m(Y_t^{A,(k)}, T^{(k)}, n_m) = \sum_{n=1}^{N} 100 w_n E'_{m,n}$

weight method, a widely used objective weight calculation method. The basic process of entropy method is shown in step 2 to step 5 in Algorithm I, which uses information entropy related to dispersion degree of $E_m(Y_t^{A,(k)}, T^{(k)}, n_m)$ to determine weights of i-indexes [34]. When deciding weights of i-indexes, $E_{m,n}$ shown in the algorithm need to be replaced by $E_m(Y_t^{A,(k)}, T^{(k)}, n_m)$.

$$I^{(k)} = \sum_{m=1}^{M} w_m E_m\left(Y_t^{A,(k)}, T^{(k)}, n_m\right) \tag{10}$$

D. The Agent Self-evolution

In the PLASE system, the process of "Management and Control" aims at gaining highest $I^{(k)}$ through controlling and adjusting $U_t^{A,(k)}$ and $U_t^{R,(k)}$, as shown in (11).

$$\left\{\hat{U}_t^{A,(k)}, \hat{U}_t^{R,(k)}\right\} = \max_{U_t^{A,(k)}, U_t^{R,(k)}} \left\{I^{(k)}\right\} \tag{11}$$

Since $U_t^{A,(k)}$ is provided by $T^{(k)}$, the process of maximizing $I^{(k)}$ through controlling $T^{(k)}$ could be expressed as (12).

$$\left\{\hat{T}^{(k)}\right\} = \max_{T^{(k)}} \left\{I^{(k)}\right\} \tag{12}$$

Control of $T^{(k)}$ could be achieved by adjusting hyperparameters $p^{(k)}$ of the power grid corrective control agent. Thus, the basic framework of the power grid corrective control AI self-evolution based on intelligence assessment in the PLASE system could be expressed by (13).

$$\{\hat{p}^{(k)}\} = \max_{p^{(k)}} \{I^{(k)}\} \tag{13}$$

To achieve self-evolution of $T^{(k)}$ by adjusting $p^{(k)}$ as shown in (10), the PLASE system turns to BO and AutoRL. AutoRL is responsible for generation, training, evaluation, and evolution of $T^{(k)}$. BO is responsible for the specific evolutionary process. The optimization objective of BO is to improve $I^{(k)}$. The Gaussian Process (GP) is taken as the probabilistic surrogate model. $p^{(k)}$ are adjusted during the evolution. These $p^{(k)}$ adjusted during evolution refer to agent neural network architectures and related hyperparameters, such as number of neural network layers, hidden size, batch size, and learning rate. Evolution of power grid corrective control AI agent is reflected by improvement of comprehensive intelligence level. The power grid corrective control AI self-evolution process in the PLASE system is shown in Fig. 5.

The self-evolution algorithm of the PLASE system, for the power grid corrective control AI agent, is shown in Algorithm 2. The power grid corrective control agent hyperparameters $\hat{p}$ for the next iteration could be obtained by optimizing acquisition function $\alpha(p)$, $\hat{p} = \arg\min_p \alpha(p)$. For the power grid corrective control AI agent constructed in this article, the Expected Improvement (EI) strategy, expressed by (14), is taken as the acquisition function for BO, which would find $p^{(k+1)}$ with low evaluation costs.

$$\begin{aligned}\alpha(p) = (\mu(I_p) - I^*)\Phi\left(\frac{\mu(I_p) - I^*}{\sigma(I_p)}\right) \\ + \sigma(I_p)\varphi\left(\frac{\mu(I_p) - I^*}{\sigma(I_p)}\right)\end{aligned} \tag{14}$$

IV. Case Study

To verify feasibility of the PLASE system for intelligence assessment and self-evolution of the power grid corrective control AI, we conduct the following experiments.

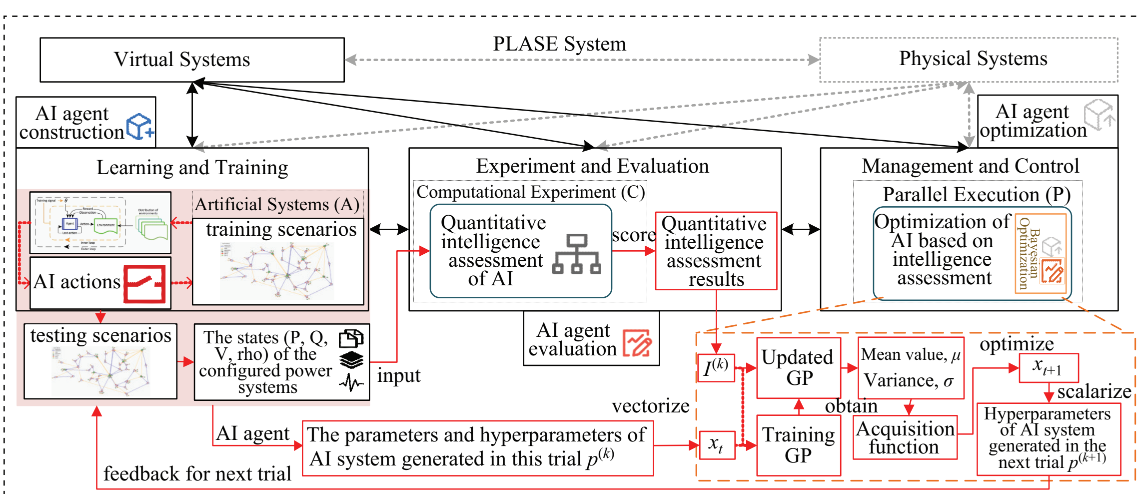

Fig. 5. The agent evolution process of the PLASE system.

Algorithm 2: Quantitative intelligence assessment based self-evolution algorithm*

1 **for** $k \leq K$ **do**
2 **if** $k = 1$ **then**
3   Choosing model parameters and hyperparameters $p^{(k)}$ randomly
4 **else**
5   Getting model parameters and hyperparameters $p^{(k)}$ from previous iteration
6 **end**
7 Constructing agent $T^{(k)}$ from $p^{(k)}$
8 **for** $s \leq S$ **do**
9   **if** *not done* **then**
10    $T^{(k)}$ provides agent actions from previous observations $Y_{t-1}^{A,(k)}$
11   Getting reward J and current observations $Y_{t-1}^{A,(k)}$
12   Training $T^{(k)}$ with reward function J
13   **end**
14   $s = s + 1$
15 **end**
16 Testing agent $T^{(k)}$ in testing scenarios n_m and getting observations $Y_{t-1}^{A,(k)}$
17 Calculating index scores in each scenario $E_{m,n}$ using scoring functions
18 Normalizing $E_{m,n}$ as $E'_{m,n}$ using Z-score method $E''_{m,n} = (E_{m,n} - \mu(E_{m,n}))/\sigma(E_{m,n})$ and Logistic method $E'_{m,n} = \frac{1}{1+e^{-E''_{m,n}}}$
19 Getting the proportion r_{mn} for normalized index scores $E'_{m,n}$ by $r_{mn} = E'_{m,n}/\sum_{m=1}^{M} E'_{m,n}$
20 Calculating information entropy $e_n = \frac{-1}{\ln(M)} \sum_{m=1}^{M} r_{mn} \ln(r_{mn})$
21 Getting weights $w_n = (1 - e_n)/\sum_{n=1}^{N}(1 - e_n)$
22 Getting total index scores $E_m(Y_t^{A,(k)}, T^{(k)}, n_m) = \sum_{n=1}^{N} 100 w_n E'_{m,n}$
23 Getting index weight w_m by repeating step 18 to step 22 or setting $w_m = 1/M$
24 Calculating intelligence level $I^{(k)} = \sum_{m=1}^{M} w_m E_m(Y_t^{A,(k)}, T^{(k)}, n_m)$
25 Saving $(T^{(k)}, I^{(k)})$ to search history $His = \{(T^{(1:k)}, I^{(1:k)})\}$
26 **if** $I^{(k)} > I^*$ **then**
27   $I^* = I^{(k)}$ as optimal intelligence assessment result so far and $T^{(k)}$ as corresponding agent
28 **else**
29   I^* as optimal intelligence assessment result so far and T^* as corresponding agent
30 **end**
31 Updating Gaussian Process (GP) with $p^{(k)}$ and $I^{(k)}$
32 Getting mean value $\mu(I_p)$ and variance $\sigma(I_p)$ from updated GP
33 Getting $p^{(k+1)}$ by optimizing $\alpha(p)$ with $\mu(I_p)$ and $\sigma(I_p)$
34 $k = k + 1$
35 **end**
36 **Return** I^* as highest intelligence level and the corresponding T^* as best agent

A. Experiment Designs

Construction and training methods of the power grid corrective control agent generated and evolved by the PLASE system in experiments are shown in the above section, which makes use of DRL technology, as well as DDDQN network. This power grid corrective control AI agent runs on the Grid2Op platform, and corresponding scenarios are also constructed on this platform [35]. In addition, it is necessary to define the following settings according to the agent evolution process in the PLASE system: target power system, i-index selected for intelligence assessment, testing scenarios for the i-index, and other experiment information.

1) The Target Power System

For the power grid corrective control problem studied in this article, target power grid is shown in Fig. 6. This is a 36-bus system extracted from the IEEE 118 bus system, including 22 generators, 37 loads, and 59 power lines. In this power system, corrective control contains 70,000 possible actions [36].

Scenario dataset of target power system has 600 different operating environments. Each environment contains operation data of power devices in five minute resolutions over a week. Operation states of target power system in one day are represented in Fig. 7. These figures take power grid operation status on weekdays (day 4 and 25) and weekends (day 13) as examples to draw typical active power of 7 loads. These operation states will be modified and improved by the power grid corrective control agent.

2) The Selection of I-indexes

Following the basic process of quantitative intelligence assessment, proper i-indexes should be selected according

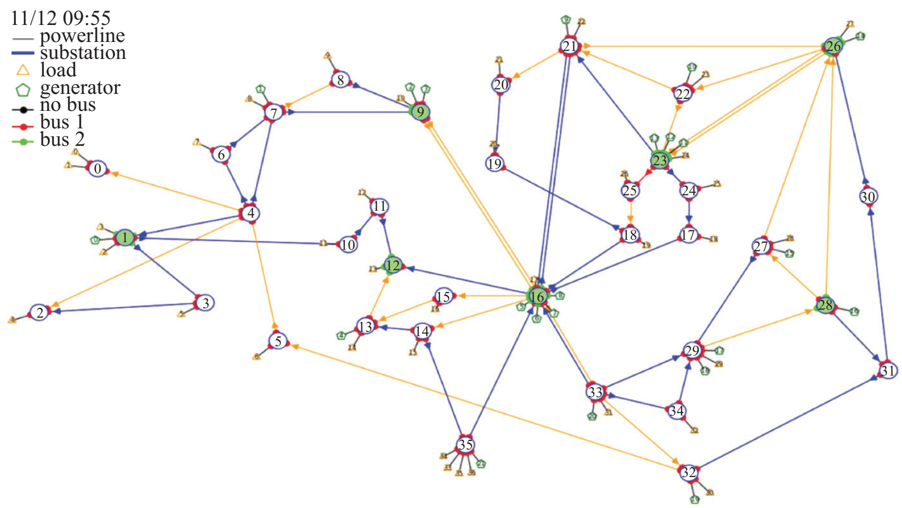

Fig. 6. The target power system.

to design goal and capabilities of the power grid corrective control AI agent.

For the power grid corrective control problem, AI agent shall control transmission of electric energy while ensuring economy and robustness of the configured power grid. The corresponding power grid control actions include, but are not limited to, network topology adjustment, power generation energy re-dispatching, and load-shedding. These actions will guarantee safe and stable operation, as well as high-quality power transmission of power grid. Thus, agent should be able to ensure stable operation of power system with minimum operational cost in any scenario and to transfer among different scenarios. Therefore, "Transferring Index" is chosen as i-index m for intelligence assessment, from intelligence assessment index system.

Typically, for the power grid corrective control AI agent used in the experiment, clearest and the simplest i-index is "Transferring index". This i-index could fulfill preliminary verification of the PLASE system.

3) The Design of Computational Experiments

After selecting i-index, we need to set appropriate testing scenarios, accordingly, considering specific index scenario requirements, similarity between training and testing scenarios, and difficulty of testing scenarios. Training and testing scenarios are set as follows:

1) Training scenarios: experiments take operation environments provided by the L2RPN competition as training scenarios. This dataset contains 600 different scenarios, covering various abnormal operation states the target power grid may encounter. Each scenario contains data in a week, which records power system operation states every five minutes. In training scenarios, target power grid encounters one of the abnormal operation states that may occur in actual operation, once a day, in the simulation. Time duration of each fault is generated randomly.

2) Testing scenarios of Transferring Index n_m: operating environments in testing scenario dataset of the L2RPN competition are selected as testing scenarios. This testing scenario dataset contains 24 independent and different scenarios. Except the absence of these scenarios in the training dataset, remaining settings and power grid operation states recording modes of testing dataset are the same as those of training dataset.

4) Other Experiment Information

Aside from designs mentioned above, the following settings also need to be defined, including adjusted hyperparameters of the PLASE system, training epochs of the power grid corrective control agent in each iteration, etc.

According to the power grid corrective control agent structure shown in Fig. 1 and training mode of the DDDQN network, hyperparameters that can be adjusted include learning rate (LR), learning rate decay rate $(LRdr)$, learning rate decay step $(LRds)$, batch size(bs), and hidden size(hs). We randomly selected three of the five hyperparameters for hyperparameter adjustment experiments, which are bs, hs, and LR. Historical time accumulation ATC will be taken as reward function for agent training. To show intelligence changes of agents with increase of their training steps more clearly, all power grid corrective control agents will be trained in 2,000 epochs and will run 864 timesteps in each epoch. Under this circumstance, the agent will train approximately 1,000,000 timesteps in total. Therefore, if total training timesteps is far less than 1,000,000, agent training fails. During intelligence assessment, 24 scenarios will be tested for each index, and timesteps of each testing scenario are 2016.

According to objectives and constraints of the corrective control problem shown in (1) and (2), the power grid corrective control agent should be able to decrease operational cost of power grid while ensuring maximum transmission power. That is to say, when controlling target power grid, operational cost of the agent should be minimized. Therefore, experiments take the operational cost-scoring function of the grid corrective control agent in testing scenarios as intelligence score of Transferring index [34]. Since only one scoring function is chosen in experiments, normalized mean value of $E_{m,n}$ in 24 testing scenarios would be intelligence score $E_m(Y_t^{A,(k)}, T^{(k)}, n_m)$ of Transferring index.

Scoring function of operational cost contains two steps.

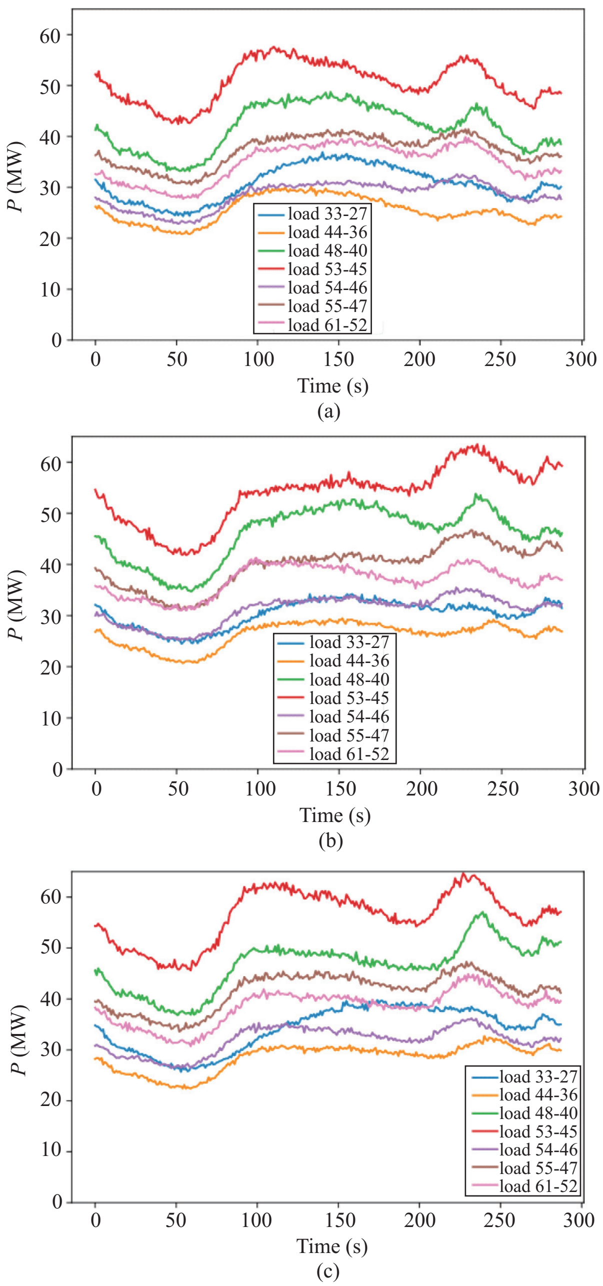

Fig. 7. The load active power of the target power system on day 4, 13, and 25. (a) Day 4. (b) Day 13. (c) Day 25.

Step 1: Calculating operational cost $C_{\text{op}}(n_m)$ in testing scenarios by (15).

$$C_{\text{op}}(n_m) = \sum_{t=0}^{t_{\text{end}}} \left(c_{\text{loss}}\left(Y_{t,P,\text{line}}^{A,(k)}, t\right) + c_{\text{Pre}}\left(Y_{t,\text{Pre}}^{A,(k)}, t\right)\right) + \sum_{t=t_{\text{end}}}^{t_{\text{E}}} c_{\text{b}}\left(Y_{t,P}^{A,(k)}, t\right) \tag{15}$$

Step 2: Converting $C_{\text{op}}(n_m)$ to $E_{m,n}$ by interpolation method. Therefore, the lower the operational cost is, the higher the $E_{m,n}$ will be.

In this way, for $T^{(k)}$ with higher scoring result, its power grid controlling effect would be better, and corresponding cost would be lower. The ability of the agent to maintain energy transmission with lower operational costs would be stronger. Thus, operational cost scoring could reflect the power grid corrective control agent's ability to control target power system intuitively.

Since only the intelligence of the Transferring index is considered in the experiments, whose weight is $w_m = 1$. Thus, according to (10), intelligence score of the Transferring index is equal to intelligence score of the agent, that is, the comprehensive intelligence level of the power grid corrective control AI agent: $I^{(k)} = E_m(Y_t^{A,(k)}, T^{(k)}, n_m)$.

To ensure randomness of scene extraction during training, guarantee credibility of experiment results, and increase persuasion of the experiment results, we conduct 10 Monte Carlo (MC) experiments for each agent with different hyperparameters. In each MC experiment, the power grid corrective control agent will be trained with the same epochs, and the intelligence level of the agent will be assessed every 15,000 timesteps. We take average result of a set of MC experiments as intelligence assessment result of the power grid corrective control agent and draw curves of the intelligence score changes with increase of training timesteps. In the intelligence score curves, x-axis refers to training timesteps of the agent in every 10^4 steps, and y-axis refers to AI intelligence score. These results will be used for the following statistical analysis.

B. The Intelligence Assessment Results of Agents

To verify the proposed intelligence assessment method can quantitatively evaluate the intelligence level of power grid corrective control agents in the first place, we take an agent with basic hyperparameters as the baseline. On this basis, we also discuss influence of hyperparameters $p^{(k)}$ on the comprehensive intelligence level $I^{(k)}$ of power grid corrective control agents, which will further verify the proposed quantitative intelligence assessment method and provide a foundation for agent evolution in the PLASE system. The hyperparameter adjusted and their value range is shown in Table I. Value of the adjusted hyperparameter is obtained through random sampling within a certain range of the exponent, and 5 different samples are chosen.

TABLE I
THE VALUE RANGE OF EACH HYPERPARAMETER

Hyperparameter	LR	bs	hs
Value range	10^{-1}–10^{-6}	2^3–2^9	2^3–2^9

1) The Results of the Baseline Agent

Hyperparameters shown in Table II will be taken as hyperparameters $p^{(k)}$ for baseline agent which will be built based on network architecture shown in Fig. 2 [8]. The power grid corrective control agent intelligence level changes with increase of training steps are shown in Fig. 8.

It can be seen in Fig. 8 the power grid corrective control agent intelligence score can be improved with increase of training steps. According to the index scoring equation used in the experiments, this agent could control power transmission of the target power system with a relatively lower operational

TABLE II
HYPERPARAMETERS FOR THE BASELINE POWER GRID CORRECTIVE CONTROL AGENT

Hyperparameter	LR	bs	hs	$LRds$	$LRdr$
Value	10^{-5}	2^6	2^7	32768	0.95

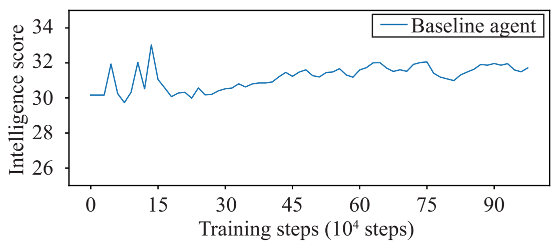

Fig. 8. The intelligence score changes of the baseline power grid corrective control agent.

cost. Ability of the agent to control the target power system improve with increase of the agent's intelligence level. The power grid corrective control effects of the agent, with lower operational cost, also improve.

The changing trend of AI intelligence score is similar to that of the reward in Fig. 9. With increase of reward, the intelligence of the agent would generally improve. Therefore, intelligence scores gained by this quantitative intelligence assessment method match intelligence changes of the agent. The proposed intelligence assessment method can quantitatively assess the intelligence level of power grid corrective control agents.

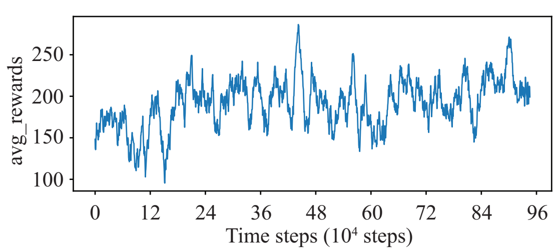

Fig. 9. The reward of the baseline power grid corrective control agent.

After intelligence assessment of the baseline agent, we explore the influence of hyperparameters on agent intelligence by assessing and comparing intelligence of the power grid corrective control agent with different hyperparameters. In this way, we verify the proposed intelligence assessment method could evaluate the intelligence level of power grid corrective control agents with different hyperparameters quantitatively, while providing a foundation for the evolution of the agent in the PLASE system.

To observe the influences of different $p^{(k)}$ on $I^{(k)}$ more intuitively, in addition to drawing intelligence score curves of power grid corrective control agents, we also make further statistical analysis of experiment results. In these statistical analyses, we gain average and variance of test scores, obtain slope and convergence steps of test score curves, and record final intelligence scores. As for the power grid corrective control agent, average of test scores represents average intelligence level; variance of test scores refers to fluctuations of the agent intelligence; slope of test score curves represents improving rate of agent intelligence; convergence steps of test score curves represent agent training and learning speed, meaning timesteps for the agent intelligence reaching stability in training; and final intelligence score represents comprehensive intelligence level $I^{(k)}$ of the agent.

2) The Results of Batch Size Adjustment

Batch size of the power grid corrective control agent refers to number of samples taken by agent during a training step. Experiment result of batch size adjustment is shown in Fig. 10, in which each curve represents the power grid corrective control AI agent with the marked batch size while the "baseline" refers to baseline agent. Statistical results are shown in Table III.

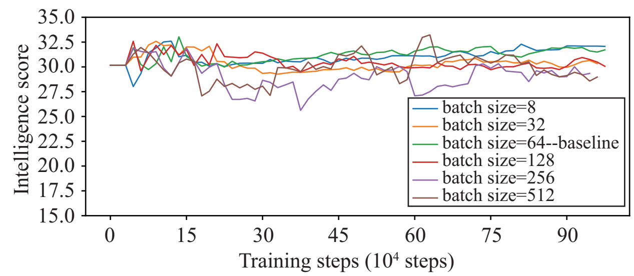

Fig. 10. The influence of the batch size on the intelligence score of the power grid corrective control agent.

TABLE III
THE STATISTICAL RESULTS OF THE BATCH SIZE ADJUSTMENT EXPERIMENT

bs	Average	Variance	Slope	Convergence steps (10^4 steps)	Final intelligence score
8	31.026	0.621	0.026	24	32.063
32	30.364	0.642	−0.012	31.5	30.331
64	31.178	0.464	0.023	24	31.717
128	30.488	0.471	−0.017	49.5	30.063
256	29.055	1.879	0.012	70.5	29.341
512	29.945	1.723	0.007	91.5	28.973

According to Fig. 10 and convergence steps in Table III, batch size mainly affects convergence speed of AI intelligence. When batch size is small, convergence speed of intelligence scores will be high, and number of convergence steps will be small. When batch size is large, intelligence score convergence speed will be much lower. With batch size increasing, convergence speed tends to slow down with some fluctuations.

3) The Results of the Hidden Size Adjustment

The hidden size of the power grid corrective control agent refers to size of the neural network. Experiment result of hidden size adjustment is shown in Fig. 11, in which each curve represents agent with the marked hidden size while the "baseline" refers to baseline agent. Statistical results are shown in Table IV.

According to Fig. 11 and average and final score in Table IV, hidden size mainly affects the initial value, average value, and final value of the intelligence score. With increase of the hidden size, the initial intelligence score gradually increases from 12 to 30 and then decreases to about 15, average

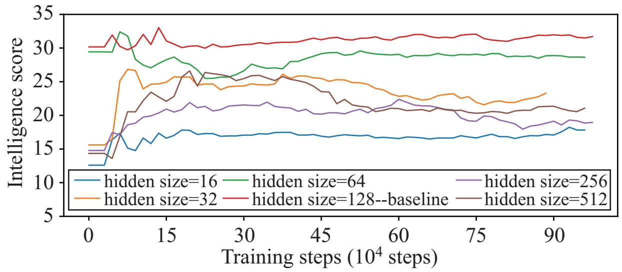

Fig. 11. The influence of the hidden size on the intelligence score of the power grid corrective control agent.

TABLE IV
THE STATISTICAL RESULTS OF THE HIDDEN SIZE ADJUSTMENT EXPERIMENT

hs	Average	Variance	Slope	Convergence steps (10^4 steps)	Final intelligence score
16	16.745	1.13	0.024	28.5	17.810
32	23.523	6.091	−0.003	33	23.310
64	28.453	1.566	0.014	57	28.636
128	31.178	0.464	0.023	24	31.717
256	20.000	2.711	0.005	33	18.943
512	21.916	8.776	−0.013	60	21.067

increases from 16 to 30 and decreases to about 20, while final score increases from 17 to 31 and decreases to about 20. These effects could be unified as the effect on the initial value of the intelligence score.

4) The Results of the Learning Rate Adjustment

Learning rate of the power grid corrective control agent refers to efficiency of AI training and learning. Experiment result of learning rate adjustment is shown in Fig. 12, in which each curve represents the power grid corrective control AI agent with the marked LR while the "baseline" refers to baseline agent. Statistical results are shown in Tables V and VI.

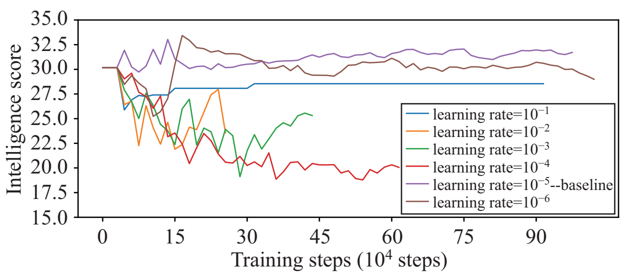

Fig. 12. The influence of the learning rate on the intelligence score of the power grid corrective control agent.

Typically, when analyzing experimental results, it is difficult to draw obvious conclusions using the original statistical results. It can be seen in Fig. 12 the agent with different LR would have different training timesteps, which is rather obvious. Hence, we add agent training steps to the statistical analysis in Table VI. Agent training steps represent duration of the agent's learning and training. It would reflect whether agent can complete the predetermined learning period and acquire knowledge.

We can also notice in Fig. 12 when $LR = 10^{-1}$, intelligence level of the power grid corrective control agent drops dramatically in the first few training steps, and its training collapses directly in 300,000 steps. Its intelligence score, after training collapse, remains the same, and its final training steps reach 900,000 steps. Thus, the intelligence score of the power grid corrective control agent with $LR = 10^{-1}$ can be classified as abnormal and will not be analyzed in the following.

TABLE V
THE STATISTICAL RESULTS OF THE LEARNING RATE ADJUSTMENT EXPERIMENT-PART I

LR	Average	Variance	Slope
10^{-1}	28.371	0.414	0.010
10^{-2}	25.573	7.750	−0.233
10^{-3}	24.816	6.973	−0.196
10^{-4}	22.358	12.619	−0.254
10^{-5}	31.178	0.464	0.023
10^{-6}	30.179	1.668	0.006

TABLE VI
THE STATISTICAL RESULTS OF THE LEARNING RATE ADJUSTMENT EXPERIMENT-PART II

LR	Convergence steps (10^4 steps)	Final intelligence score	Training steps (10^4 steps)
10^{-1}	21	28.536	90
10^{-2}	∞	23.571	24
10^{-3}	∞	25.310	42
10^{-4}	46.5	20.060	60
10^{-5}	24	31.717	96
10^{-6}	27	28.994	100.5

According to Fig. 12 and the average, final intelligence score, and training steps in Tables V and VI, learning rate mainly affects whether the power grid corrective control agent can acquire knowledge, and has a certain impact on the comprehensive intelligence level of the power grid corrective control agent. When LR is high, the intelligence score experiences a cliff fall, and agent training steps are extremely short. There are more "illegal" and "ambiguous" actions generated by these agents during training, as recorded in the PLASE system operation data. Thus, it is difficult for these agents to complete the pre-determined 1,000,000 timesteps training period and acquire knowledge. With decrease of LR, the intelligence score and training steps gradually increases. The "illegal" and "ambiguous" actions generated by these agents gradually reduce. These agents can gradually complete their predetermined training period and acquire knowledge. Therefore, the smaller the LR, the more likely the agent is to acquire knowledge and improve its intelligence level. However, if the LR is too small, the intelligence level of the power grid corrective control agent may become lower in return.

In conclusion, the proposed intelligence assessment method could assess the intelligence level of the power grid corrective control agent with different hyperparameters, and adjusting hyperparameters of power grid corrective control agents could affect their intelligence level. Specifically, batch size affects convergence speed of the power grid corrective control agent intelligence in training, hidden size affects initial intelligence of the agent, and learning rate affects whether agent can acquire knowledge, as well as final intelligence level of the agent. By adjusting hyperparameters, it is possible to improve the intelligence level of the power grid corrective control agent and promote the agent evolution.

C. The Best Power Grid Corrective Control Agent Generated by the PLASE System Through Self-evolution

From the experiments mentioned above, we have made clear the influence of different hyperparameters on the power grid corrective control agent intelligence. It is verified the power grid corrective control agent intelligence level can be changed by adjusting its hyperparameters. On this basis, we can use the PLASE system to carry out agent evolution experiment.

Evolution of the power grid corrective control agent aims to improve its intelligence level, in order to improve power grid operation states and decrease system operational cost. In general, the power grid corrective control agent evolution goal can be summarized as three points: 1) increase of final intelligence score, 2) decrease of convergence steps of intelligence score curve, and 3) increase of slope of the intelligence score curve. Among these goals, increase of the intelligence score, meaning improvement of the comprehensive intelligence level of the agent, is the primary goal, which accounts for a large proportion of evolution.

When introducing the method and operation process of the PLASE system, intelligence score improvement is taken as an example of the power grid corrective control agent evolution. In practice, however, the evolutionary goal of the PLASE system can be replaced according to specific manifestation of intelligence improvement. Therefore, in this section, we take improvement of final intelligence score, reduction of convergence steps, and improvement of slope as evolution goals of the power grid corrective control agent generated in the PLASE system. Among these goals, improvement of the final intelligence score has a higher proportion. The power grid corrective control agent evolution experiment will be conducted accordingly.

We briefly draw the 3-dimensional distribution curve of the power grid corrective control agent self-evolution accomplished by the PLASE system, as shown in Fig. 13. In Fig. 13, slope of the intelligence score curve is taken as x-axis, convergence steps of the intelligence score curve are taken as y-axis, and final intelligence score is taken as z-axis.

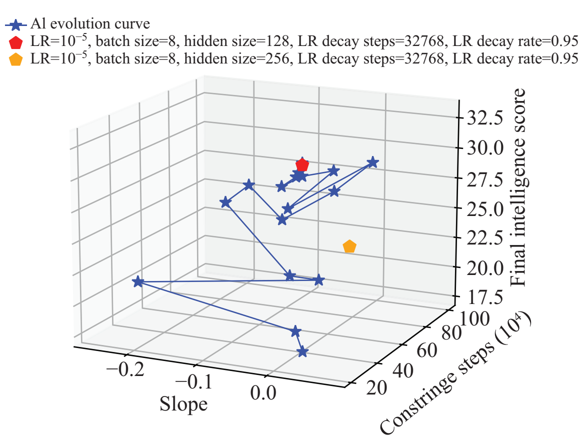

Fig. 13. The 3-dimensional distribution curve of the agent evolution.

The point, with 1,000,000 convergence steps, represents the situation where convergence steps of the agent intelligence score are ∞ or intelligence score is not convergent. The blue line represents agent evolution curve, and points on the blue line represent the power grid corrective control agents with different hyperparameters. Red and orange points, respectively, represent the two best power grid corrective control agents with different hyperparameters fitted by BO. These best agents have either the relatively highest intelligence score, lowest convergence steps, or biggest slope.

To further analyze intelligence level changes of the best agent represented by the red point and orange point in Fig. 13, and clarify the real best agent, we draw intelligence change curves and analyze statistical results of the two best agents. Hyperparameters p^* of the best power grid corrective control agent T^*, obtained by evolution, are shown in Table VII. These hyperparameters will be used in the following MC experiments.

TABLE VII
THE POSSIBLE HYPERPARAMETERS OF THE BEST POWER GRID CORRECTIVE CONTROL AGENT

Agent	LR	bs	hs	$LRds$	$LRdr$
The agent showed by the red point	10^{-5}	8	128	32768	0.95
The agent showed by the orange point	10^{-5}	8	256	32768	0.95

Figures 14 and 15 show average and best result of the Monte Carlo experiment of the power grid corrective control agent shown by the red point in Fig. 13, with hyperparameters of $LR = 10^{-5}$, $bs = 8$, $hs = 128$, $LRds = 32768$, and $LRdr = 0.95$, respectively. Their statistical results are shown in Table VIII. Figs. 16 and 17 show average and best result of the Monte Carlo experiment of the power grid corrective control agent shown by the orange point in Fig. 13, with hyperparameters of $LR = 10^{-5}$, $bs = 8$, $hs = 256$, $LRds = 32768$, and $LRdr = 0.95$, respectively. Their statistical results are shown in Table IX.

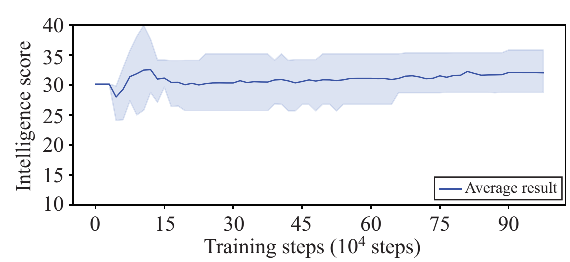

Fig. 14. The result of the agent shown by the red point.

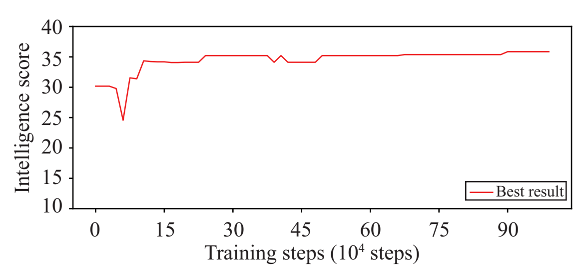

Fig. 15. The best result of the agent showed by the red point.

TABLE VIII
THE STATISTICAL RESULTS OF THE AGENT SHOWED BY THE RED POINT

Average	Variance	Slope	Convergence steps (10^4 steps)	Final intelligence score	Training steps (10^4 steps)
31.668	0.841	0.031	24	32.830	98

TABLE IX
THE STATISTICAL RESULTS OF THE AGENT SHOWED BY THE ORANGE POINT

Average	Variance	Slope	Convergence steps (10^4 steps)	Final intelligence score	Training steps (10^4 steps)
26.307	10.117	0.082	31.5	27.345	96

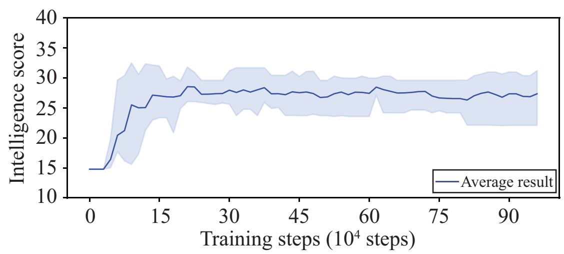

Fig. 16. The result of the agent shown by the orange point.

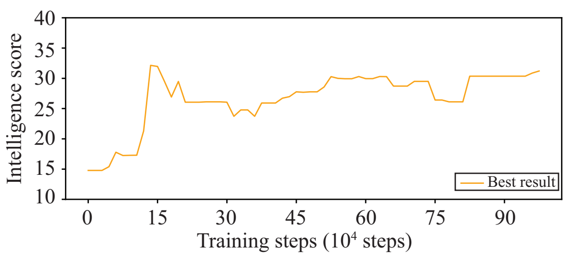

Fig. 17. The best result of the agent showed by the orange point.

It can be seen from the aforementioned results the intelligence level of the power grid corrective control agent with $LR = 10^{-5}$, $bs = 8$, $hs = 128$, $LRds = 32768$, and $LRdr = 0.95$ increases significantly in training. Fluctuation of intelligence score is small, and training is generally stable. The final comprehensive intelligence level of the power grid corrective control agent is relatively high ($I^* = 32.830$), increasing by approximately 88% compared to intelligence of the original agent. Intelligence level of the power grid corrective control agent with $LR = 10^{-5}$, $bs = 8$, $hs = 256$, $LRds = 32768$, and $LRdr = 0.95$ also increases significantly. Intelligence score fluctuates greatly, while training is comparatively stable. Comprehensive intelligence level of the power grid corrective control agent ($I^* = 27.345$) is lower than of the previous one, which still increases approximately by 59% compared to original agent.

Since improvement of the final intelligence score is much more important for the PLASE system in this case, according to experiment results, hyperparameters of the best power grid corrective control agent p^* are: $LR = 10^{-5}$, $bs = 8$, $hs = 128$, $LRds = 32768$, and $LRdr = 0.95$. The best comprehensive intelligence level I^* of the power grid corrective control agent can reach 35.856. Intelligence score of the agent is calculated by operational cost of the agent. The higher the intelligence score is, the lower the agent's operational cost will be. According to the goal and constraints of the power grid corrective control problem, the agent would ensure power system robustness with a lower operational cost. Thus, the best power grid corrective control agent generated by the PLASE system could guarantee robustness of the target power grid with a much lower operational cost. To sum up, by using the PLASE system and BO algorithm, the power grid corrective control agent self-evolution guided by intelligence improvement can be realized.

V. CONCLUSION

This article proposes the PLASE system based on the computational framework of parallel systems, to achieve self-evolution of the power grid corrective control AI agents guided by their intelligence level improvement. When constructing the PLASE system, this article resorts to the ACP approach, establishes a quantitative intelligence assessment index system, and frames an AI self-evolution method by introducing the BO and AutoRL into the system. Experiments confirm the PLASE system can objectively and quantitatively assess the intelligence level of the power grid corrective control agent AI. Experiments also reveal the PLASE system can result in power grid corrective control AI evolution based on the intelligence assessment results, providing improvement of the intelligence level. The intelligence level of the best agent generated by the PLASE system increases by approximately 88%. AI evolution in experiments is realized by adjusting hyperparameters of the power grid corrective control agent.

In conclusion, the PLASE system could generate the power grid corrective control AI agent with the relatively best intelligence. In AI evolution, human intervention would be limited, thus benefiting power engineers who are not well-equipped with AI knowledge. Furthermore, according to the intelligence assessment index system, the PLASE system could apply to other AI agents.

In following research, multiple intelligence indexes and different index scoring equations will be chosen, and different parameters and hyperparameters will be selected to adjust, to explore and realize quantitative intelligence assessment and self-evolution of other AI and hybrid intelligence systems for different power grid control tasks. In the meantime, the operating process of the PLASE system will remain the same. Other parameters and hyperparameters that might be adjusted include type of the neural network, layer number of the neural network, number of cells in each layer, learning rate decay rate and steps, type of activation function and optimizer of the reinforcement learning, number of the samples related to human knowledge, number of preserved control policies and schemes given by humans, number of generated samples, sampling probability of sample data, capacity of experience pool, capacity of training data given by humans, etc. Especially, the last 6 parameters usually belong to the hybrid intelligence systems.

REFERENCES

[1] F. Chollet. (2019 Nov.). On the measure of intelligence [Online].

Available: https://arxiv.org/abs/1911.01547.

[2] A. Boggust, B. Hoover, A. Satyanarayan, and H. Strobelt, "Shared interest: measuring human-AI alignment to identify recurring patterns in model behavior," in *Proceedings of 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 10, doi: 10.1145/3491102.3501965.

[3] J. Parker-Holder, R. Rajan, X. Y. Song, A. Biedenkapp, Y. J. Miao, T. Eimer, B. H. Zhang, V. Nguyen, R. Calandra, A. Faust, F. Hutter, and M. Lindauer, "Automated reinforcement learning (AutoRL): A survey and open problems," *Journal of Artificial Intelligence Research*, vol. 74, pp. 517–568, Jun. 2007.

[4] J. D. Morales, P. N. Papadopoulos, and J. V. Milanović, "Feasibility of different corrective control options for the improvement of transient stability," in *2017 IEEE Manchester PowerTech*, 2017, pp. 1–6, doi: 10.1109/PTC.2017.7981015.

[5] R. Moreno, D. Pudjianto, and G. Strbac, "Transmission network investment with probabilistic security and corrective control," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3935–3944, Nov. 2013, doi: 10.1109/TPWRS.2013.2257885.

[6] L. Roald, S. Misra, T. Krause, and G. Andersson, "Corrective control to handle forecast uncertainty: a chance constrained optimal power flow," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1626–1637, Mar. 2017, doi: 10.1109/TPWRS.2016.2602805.

[7] A. Kelly, A. O'Sullivan, P. de Mars, and A. Marot. (2020, Mar.). Reinforcement learning for electricity network operation [Online]. Available: https://arxiv.org/abs/2003.07339.

[8] P. D. Xu, J. J. Duan, J. Zhang, Y. Z. Pei, D. Shi, Z. W. Wang, X. Z. Dong, and Y. Z. Sun, "Active power correction strategies based on deep reinforcement learning—part I: a simulation-driven solution for robustness," *CSEE Journal of Power and Energy Systems*, vol. 8, no. 4, pp. 1122–1133, Jul. 2022, doi: 10.17775/CSEEJPES.2020.07090.

[9] S. Y. Chen, J. J. Duan, Y. Y. Bai, J. Zhang, D. Shi, Z. W. Wang, X. Z. Dong, and Y. Z. Sun, "Active power correction strategies based on deep reinforcement learning—part II: a distributed solution for adaptability," *CSEE Journal of Power and Energy Systems*, vol. 8, no. 4, pp. 1134–1144, Jul. 2022, doi: 10.17775/CSEEJPES.2020.07070.

[10] S. Y. Chen, Y. Y. Bai, and Z. Jun, "Dynamic load shedding strategy using distributional deep reinforcement learning in power system emergency control," in *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, 2020, pp. 248–253, doi: 10.1109/EI250167.2020.9346749.

[11] M. S. Li, H. M. Zhang, T. Y. Ji, and Q. H. Wu, "Fault identification in power network based on deep reinforcement learning", *CSEE Journal of Power and Energy Systems*, vol. 8, no. 3, pp. 721–731. May 2022, doi: 10.17775/CSEEJPES.2020.04520.

[12] J. M. L. Asensio, J. Peralta, R. Arrabales, M. G. Bedia, P. Cortez, and A. L. Peña, "Artificial intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters", *Expert Systems with Applications*, vol. 41, no. 16, pp. 7281–7290, Nov. 2014.

[13] J. Insa-Cabrera, D. L. Dowe, S. España-Cubillo, M. Hernández-Lloreda, and J. Hernández-Orallo, "Comparing humans and AI agents," in *4th International Conference on Artificial General Intelligence*, 2011, pp. 122–132.

[14] M. J. Kim, K. J. Kim, S. Kim, and A. K. Dey, "Evaluation of StarCraft artificial intelligence competition bots by experienced human players," in *Proceedings of 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 1915–1921, doi: 10.1145/2851581.2892305.

[15] T. Schaul, J. Togelius, J. Schmidhuber. (2011, Sep.). Measuring intelligence through games [Online]. Available: https://arxiv.org/abs/1109.1314.

[16] T. Y. Zhang, T. L. Gao, P. D. Xu, and J. Zhang, "A review of AI and AI intelligence assessment," in *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, 2020, pp. 3039–3044, doi: 10.1109/EI250167.2020.9347303.

[17] X. He, K. Y. Zhao, and X. W. Chu, "AutoML: a survey of the state-of-the-art", *Knowledge-Based Systems*, vol. 212, pp. 106622, Jan. 2021, doi: 10.1016/j.knosys.2020.106622.

[18] M. Feurer, A. Klein, K. Eggensperger, J. T. Springenberg, M. Blum, and F. Hutter, "Methods for improving Bayesian optimization for AutoML," presented at *Workshop and Conference Proceedings 2015*, 2015.

[19] H. F. Jin, Q. Q. Song, and X. Hu, "Auto-Keras: an efficient neural architecture search system," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1946–1956.

[20] J. Liang, E. Meyerson, B. Hodjat, D. Fink, K. Mutch, and R. Miikkulainen, "Evolutionary neural AutoML for deep learning," in *Proceedings of Genetic and Evolutionary Computation Conference*, 2019, pp. 401–409, doi: 10.1145/3321707.3321721.

[21] R. S. Olson and J. H. Moore, "TPOT: A tree-based pipeline optimization tool for automating machine learning," in *Proceedings of the Workshop on Automatic Machine Learning*, New York, USA, 2016, pp. 66–74.

[22] E. Real, C. Liang, D. R. So, and Q. V. Le, "AutoML-Zero: evolving machine learning algorithms from scratch," in *Proceedings of the 37th International Conference on Machine Learning*, 2020, pp. 742.

[23] A. M. Alaa and M. van der Schaar. "AutoPrognosis: automated clinical prognostic modeling via Bayesian optimization with structured kernel learning," in *Proceedings of the 35th International Conference on Machine Learning*, 2018, pp. 139–148.

[24] Y. T. Chen, A. Huang, Z. Y. Wang, I. Antonoglou, J. Schrittwieser, D. Silver, and N. de Freitas. (2018, Dec.). Bayesian optimization in AlphaGo [Online]. Available: https://arxiv.org/abs/1812.06855.

[25] J. M. Hernández-Lobato, M. A. Gelbart, R. P. Adams, M. W. Hoffman, and Z. Ghahramani, "A general framework for constrained Bayesian optimization using information-based search," *Journal of Machine Learning Research*, vol. 17, no. 160, pp. 1–53, Sep. 2016.

[26] J. C. Lévesque, C. Gagné, and R. Sabourin, "Bayesian hyperparameter optimization for ensemble learning," in *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*, 2016, pp. 437–446.

[27] J. C. Lévesque, A. Durand, C. Gagné, and R. Sabourin, "Bayesian optimization for conditional hyperparameter spaces," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 286–293, doi: 10.1109/IJCNN.2017.7965867.

[28] Z. Wang, C. Gehring, P. Kohli, and S. Jegelka, "Batched large-scale Bayesian optimization in high-dimensional spaces," in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, 2018, pp. 745–754.

[29] J. Zhang, P. D. Xu, and F. Y. Wang, "Parallel systems and digital twins: a data-driven mathematical representation and computational framework," *Acta Automatica Sinica*, vol. 46, no. 7, pp. 1346–1356, Jul. 2020.

[30] P. I. Frazier. (2018, Jul.). A tutorial on Bayesian optimization [Online]. Available: https://arxiv.org/abs/1807.02811.

[31] H. Morgan, "An analysis of Gardner's theory of multiple intelligence," *Roeper Review*, vol. 18, no. 4, pp. 263–269, Feb. 1996, doi: 10.1080/02783199609553756.

[32] Y. F. Feis, "Raven's progressive matrices," in *Encyclopedia of Cross-Cultural School Psychology*, C. S. Clauss-Ehlers, Ed. New York: Springer, 2010, pp. 787, doi: 10.1007/978-0-387-71799-9_344.

[33] S. S. Cai and X. D. Xue, "Artificial intelligence and human intelligence: a view on human-computer competition from the five-level theory of cognitive science," *Journal of Peking University (Philosophy and Social Sciences)*, vol. 53, no. 4, pp. 145–154, Jul. 2016.

[34] W. Lin and L. B. Shao. (2013, Dec.). Evaluation on the financial competitiveness of Chinese listed real estate companies based on entropy method [Online]. Available: https://arxiv.org/abs/1302.2493.

[35] RTE-France. (2021). Grid2Op [Online]. Available: https://github.com/rte-france/Grid2Op.

[36] A. Marot, B. Donnot, G. Dulac-Arnold, A. Kelly, A. O'Sullivan, J. Viebahn, M. Awad, I. Guyon, P. Panciatici, and C. Romero, "Learning to run a power network challenge: a retrospective analysis," in *Proceedings of NeurIPS 2020 Competition and Demonstration Track*, 2021, pp. 112–132.

Tianyun Zhang (S'20) received the B.S. degree in Electrical Engineering and Automation from Wuhan University, Wuhan, China, in 2016 and the M.S. degree in Electrical Engineering from the University of Southern California, USA, in 2018. She is currently pursuing the Ph.D. degree in Electrical Engineering at the School of Electrical Engineering and Automation, Wuhan University, China. Her research interest includes hybrid intelligence in power system operation and control.

Jun Zhang (M'09–SM'14) received his B.E. and M.E. degrees in Electrical Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2003 and 2005, respectively, and his Ph.D. degree in Electrical Engineering from Arizona State University, U.S.A., in 2008. He is currently a Professor at the School of Electrical Engineering and Automation, Wuhan University. He authored/co-authored over 200 peer-reviewed publications. He is an Associate Editor of IEEE Transactions on Computational Social Systems, Associate Editor of Acta Automatica Sinica, and Associate Editor of IEEE/CAA Journal of Automatica Sinica. His research expertise is in the areas of signal processing, big data, artificial intelligence, Blockchain, and their applications in complex and intelligent systems.

Feiyue Wang (S'87–M'89–SM'94–F'03) received the Ph.D. degree in Computer and Systems Engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990. He founded the CAS Center for Social Computing and Parallel Management in 2008 and became the State Specially Appointed Expert and the Founding Director of the State Key Laboratory for Management and Control of Complex Systems in 2011. His research interests include parallel intelligence, social computing, and knowledge automation.

Peidong Xu (S'19) received the M.S. and Ph.D. degrees in the School of Electrical Engineering and Automation from Wuhan University, Wuhan, China, in 2018 and 2022, respectively. He is currently pursuing the postdoctoral degree in Electrical Engineering at the School of Electrical Engineering and Automation, Wuhan University, Wuhan, China. His research interests include power system operation and control, artificial intelligence, and electric vehicles. He authored/co-authored over 20 peer-reviewed publications.

Tianlu Gao (S'22) received the M.S. degree from the School of Electrical and Computer, University of Denver, USA, in 2019. He is currently pursuing the Ph.D. degree at the School of Electrical and Automation, Wuhan University, Wuhan, China. He authored/co-authored over 30 peer-reviewed publications, including 13 Science Citation Index (SCI) papers. His main research interests are the application of Simulation, AI, NLP, and Blockchain in power system operation and control.

Haoran Zhang received the B.S. degree from the University of Electronic Science and Technology of China (UESTC), China, in 2021. He is currently pursuing the M.S. degree at the School of Electrical Engineering and Automation, Wuhan University. His research interests include power grid dispatching and the applications of machine learning in power systems.

Ruiqi Si received the B.S. degree from the School of Electrical Engineering, Sichuan University, China, in 2021. She is currently pursuing the M.S. degree at the School of Electrical and Automation, Wuhan University, Wuhan, China. Her current research interests include the optimization of power systems, reinforcement learning, and multi-agent system.