IEEE Open Journal of the
Computer Society

Received 26 June 2021; accepted 13 July 2021. Date of publication 16 July 2021;
date of current version 31 August 2021. The review of this paper was arranged by Guest Editors Sahil Garg, Alagan Anpalagan, Shree Krishna Sharma, Li-Chun Wang, Muhammad Zeeshan Shakir, Ning Zhang, and Linglong Dai.

Digital Object Identifier 10.1109/OJCS.2021.3097540

DRESIA: Deep Reinforcement Learning-Enabled Gray Box Approach for Large-Scale Dynamic Cyber-Twin System Simulation

ZHOUYANG LIN [1,2], KAI LI [1], YANG YANG [1,4,5] (Fellow, IEEE), FANGLEI SUN [3], LIANTAO WU [1], PANPAN SHI [1,2], SONG CI [6] (Senior Member, IEEE), AND YONG ZUO [7]

[1] Shanghai Institute of Fog Computing Technology, School of Information Science and Technology, Shanghaitech University, Shanghai 201210, China
[2] Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China
[3] Artificial Intelligence and Digital Art Lab, School of Creativity and Art, Shanghaitech University, Shanghai 201210, China
[4] Research Center for Network Communication, Peng Cheng Laboratory, Shenzhen 518000, China
[5] Shenzhen SmartCity Technology Development Group Company Ltd., Shenzhen 518046, China
[6] Electrical Engineering Department, Tsinghua University, Beijing 100084, China
[7] College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China

CORRESPONDING AUTHOR: KAI LI (e-mail: likai@shanghaitech.edu.cn); YANG YANG (e-mail: yangyang@shanghaitech.edu.cn).

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB2104300, and in part by the National Natural Science Foundation of China Key Project Program under Grant 61932014.

ABSTRACT The massive data generated by large-scale dynamic systems makes its optimization facing a tough challenge. Traditional White Box-based methods directly model the internal operating mechanism of the system, so massive amounts of measured data need to be handled, which is costly and time-consuming. The poor interpretability of the Black Box-based methods makes it difficult to adapt to the dynamic environment. Thus we propose a novel Gray Box-based approach namely Deep Reinforcement Learning-enabled Constraint Set Inversion Algorithm (DRESIA), which establishes a quantitative model of the nonlinear interoperability effects of system internal states which simplifies the White Box's complex mechanism of reconstruction and prediction and retains the interpretability of the model, therefore improves the prediction efficiency of feasible region while also improving the generalization ability. It further improves the dynamic adaptability of the modeling environment, which provides a new performance balancing scheme for system modeling. Under the premise that the large-scale 5G Cyber-Twin system satisfies the given Quality of Service (QoS) requirements, we perform DRESIA to realize the efficient and dynamic optimal search of feasible region, the results show that the DRESIA reduces the computational cost, and balances the accuracy and robustness of the feasible region, which validate the effectiveness and superiority of Gray Box-based approach.

INDEX TERMS Cyber-twin, digital twin, dynamic system, white-box, black-box, gray-box, fuzzy measure, choquet integral, deep reinforcement learning, feasible region inversion, massive MIMO.

I. INTRODUCTION

Most of the existing large-scale system models in the vast majority of cases such as cyber-twin [12], digital twin [25], intelligent transportation systems(ITS) [44], etc. are dynamic and stochastic. For a cyber-twin technology-based simulation platform, the analytical description does not fully cover the behaviour of the actual systems. To evaluate the quality of the future system in the preliminary design stage, or improve the function of the system and the quality index value in the operation stage, a region of efficiency should be found to prove the stable service of the real system. The feasible region is defined as a set of combinations of system parameters, namely a constraint set, for which the values of the selected quality indicators are in the sense of Pareto better than the pre-selected boundary values.

As a kind of complex time-varying system, the constraint set of large-scale wireless communication system is the reference range for the parameters that ensure its steady operation. In our previous works, we have adopted both the White Box-based and Black Box-based approaches to optimize the performance indicators. The White Box-based method using Channel characteristics such as detailed ray-tracing data captured by our designed code-division multiplexing-based parallel channel sounder [38], as the planning parameters to generate the 3D channel matrix based on the 3GPP 38.901 protocol [13]. A complete simulation calculation process generates higher-level path loss, throughput and other performance indicators to obtain overall network performance. The Black Box-based method [40], [41] that we proposed training the neural network with ray-tracing data and performance indicators generated by the White-Box method, and rapid network performance evaluation is achieved by reusing the trained neural network.

In the existing studies, Black-Box-based technology [33] devised a specific model according to a specific scenario to find the global optimum, thus finding the constraint set. [4] performed the ascending iteration method with the largest gradient through obtaining the local approximation of the objective function. Therefore, this method realized the online parameter optimization of the system containing random factors but it is not ideal in a multi-parameter simulation system. To tackle this problem, a Gradient Descent-based algorithm [21] with a penalty method was proposed. It converted an optimization problem constrained by the coverage condition into a simple form with only lower and upper bound conditions, and thus a discrete coverage index was converted into a continuous coverage index. Then it can be solved by Gradient Descent Optimization. This method can be applied to various optimization conditions. However, a large amount of real-time data feedback is required to achieve the expected effect, the parameter range is also limited. Otherwise, distortion information will be generated in the process steps from discrete to continuous. And this method is not applicable to the situation where there are too many parameters and the parameters interfere with each other. Taguchi Method (TM) [10], [30], [35] as a Data-Driven technology was developed to optimize manufacturing processes and then imported into several engineering fields, including wireless networks. Then it can be used for multi-objective design optimization [28] and wireless networks parameter optimization [10] etc. However, for large-scale systems with numerous parameters, wide range of parameter values, and long simulation time, TM will cause severe time overhead.

For the defects of high data sampling rate, e.g. high storage requirement and low generalization, exhaustive methods or deep learning methods cannot be used in large-scale dynamic system, therefore, heuristic learning is currently one of the mainstream way to deal with this scenario. The Genetic Algorithm (GA) [10], [14], [14], [27] is widely used to quickly search for the optimal solution of combinatorial optimization problems such as gateway placement optimization [1] and coverage optimization of mobile wireless sensor networks [20] etc. However, the Hamming Cliff [11] makes the crossover and mutation steps in the GA hard to cross. Otherwise, for the problem to be solved in this article, parameters with different meanings cannot be expressed in a unified binary code. Moreover, it could face a local optimal dilemma sometimes. Particle Swarm Optimization (PSO) [43] and Ant Colony Optimization (ACO) [7], [8] as swarm intelligence algorithms performed a group of unintelligent or slightly intelligent individuals (agent) through cooperation to show intelligent behaviour, thus providing a new possibility for solving complex problems [32]. In this way, PSO and ACO can be used for optimizing the radio network parameters [10], wireless sensor network path optimization [45], network inference, 2-D coverage optimization in wireless sensor network [3], cellular network spectrum allocation [42], and parameter estimation [46]. These algorithms require a bulk of data set to give feedback to agents. Therefore, PSO and ACO will face difficulty solving the objective problem while the data set is not dense enough, and the converge time is too long. Simulated Annealing (SA) [8], [18] has the same high robustness as PSO. It can be used for the location distribution of non-anchor nodes in sensor networks [17], finding the shortest path between wireless network sensors [8] and wireless sensor network layout problem [23] etc. But SA has the same problem of high calculation amount and dense data set requirement.

Considering the defect of the above methods, Fuzzy Measure (FM) [16] as a special non-additive measure has good performance in the evaluation of subjective, complex systems. There are two well-known fuzzy measures, λ-measure and P-measure. The former is not a close form, and the latter is not sensitive. The non-additive model based on Choquet integral [16] has great practical use in many fields, especially in the interaction among predictive attributes towards the objective attributes. Through Choquet integral, the nonlinear relationship between the parameter and the objective output is transformed into a linear additive form. Therefore, we can characterize the mapping of simulation system parameters to system performance [37]. [15] proposed a new genetic algorithm for nonlinear multi-regression based on generalized Choquet integral with respect to signed fuzzy measures. Unlike previous work of Choquet integral, the speciality of this new model is that the interaction among predictive attributes towards the objective attribute can be properly reflected through a signed fuzzy measure and the relevant Choquet integral. It shortens the running time of the program and increases the accuracy of the result. [2], [6], [36] proposed a cross-layer design optimization algorithm based on Choquet integral and its desired effect is obtained, the amount of calculation and the memory required for calculation are greatly reduced. This method established an observable and interpretable parameter model, but does not yet have the ability to flexibly adjust parameters.

With the support of a large number of data samples, the fitted Black Box-based method solves the problem of rapid prediction, but the cost of obtaining samples and training is too high, and it cannot dynamically adapt to environmental changes. The White Box-based method does not require data sampling, but the calculation process such as reconstruction and prediction is complicated, low in efficiency and high in cost. In this article, to weigh the accuracy and efficiency of system measurement and reconstruction, we proposed a novel Gray Box method containing two procedures to solve the observability, controllability, and interpretability of system parameters. The Gray Box method adopts the Choquet integral-based Interdependency and Significance Analysis (CISA) algorithm to obtain the combinations of parameters from a limited number of data sets that have the most significant impact on the MIMO system's performance. In the case of uniform sampling of parameters within the value range, CISA can accurately reflect the contribution of different parameter combinations to the MIMO system's performance. Moreover, for practical usage scenarios, a Deep Reinforcement Learning-enabled Constraint Set Inversion Algorithm (DRESIA) as another procedure of the Gray Box method is adopted, which can find a combination of parameters that ensure the QoS qualified when individual parameters change due to the external environment. The key contributions of this work are summarized as follows.

- CISA is proposed to characterize the contribution of parameter combinations in the MIMO system, therefore significantly enhance our capability for system behaviour characterization and provide better insights for system running.
- A novel method with observable, adjustable and interpretable parameters namely DRESIA is proposed to overcome the problem of non-convergence caused by sparse data set. DRESIA adopts the principle of data fusion that can combine the result of Choquet integral with the measured data as the reward of the deep reinforcement learning-enabled constraint set inversion algorithm to find the parameter constraint set that meets the system performance requirements. And the dynamically updated data set guarantees that the DRESIA can adapt to dynamic changes of the system.

To evaluate the effectiveness of Gray Box method on dynamic stochastic system, we need a specific simulation platform. As a key enabling technology for 5G, Massive MIMO significantly increases the system capacity through Space Division Multiplexing (SDM), and meanwhile it also increases the complexity of network planning and optimization. The difficulty of 5G large-scale MIMO network planning and optimization lies in the complexity of 3D wireless channel and the need for a large number of parameters to optimize the parameter configuration of wireless system from the perspective of spatial, frequency and time domain. The parameters that directly affect the coverage effect of large-scale MIMO include antenna selection, base station deployment location, antenna downtilt, azimuth, power and other major engineering parameters, as well as other major wireless protocol parameters. The system capacity, wireless coverage, and QoS of massive MIMO strongly depend on the 3D channel propagation environment. While in the real coverage scenario, the 3D channel is normally constructed by the 3D spatial characteristics of the propagation environment. Because of these characteristics, we adopt the 3D channel model reconstructed by the ray-tracing data generated based on the 3D map as the simulation platform, and perform the Gray Box method on it to verify its feasibility.

The rest of the paper is organized as follows. In Section II, the characteristics of three different models was briefly introduced and compared. In Section III, the CISA algorithm was introduced to analyzing the contribution degree of each system parameter to the output. In Section IV, the novel approach namely DRESIA was presented for searching the configuration combinations of parameters that meet the requirements. In Section V, the simulation results was given to verify the effectiveness of the proposed approaches in Section III and Section IV. Finally, this work is concluded in Section VI.

II. SYSTEM MODEL AND COMPONENTS

The advantages of comprehensive coverage, good adaptability, and good portability make Deep Learning (DL) widely used in many fields, especially efficient for the simulation of applications that contain high-dimensional data or factors that are difficult to quantify mathematically. Therefore, the DL-based Black-Box simulation modeling system is widely used in industry to replace some White-Box functions that are time-insensitive and time-invariant.

However, the communication environment will change with climate, time and real-time service volume, etc. To maintain high-quality transmission in the communication system, except for some fixed-parameter function, other parameters should be observable for system testing. Similarly, while these parameters are observable, they should also be adjustable so that the system can be optimized. However, the Black-Box function is difficult to be observed and adjusted, which makes the system update and optimization very difficult. Furthermore, without newly observed data, the Black-Box function cannot be updated, resulting in poor generalization.

As one of the White Box-based approaches, FM-based quantitative model overcomes the shortages of unobservable and unadjustable parameters of the Black-Box system. As shown in Fig. 1, the FM-based White Box system is contributed by a set of FM-based functions, which require a certain amount of uniformly sampled data, and select the appropriate function form to fit. The system can change the appropriate function form according to different operating environments. Furthermore, the FM-based function's parameters are observable and adjustable since it builds a quantification table. Although the parameters are adjustable, the model cannot evolve because data sampled in the same period time could

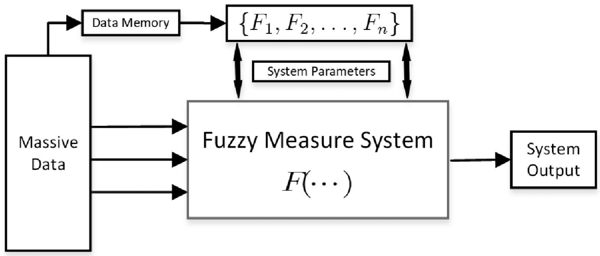


FIGURE 1. **Fuzzy Measure based White Box system.**

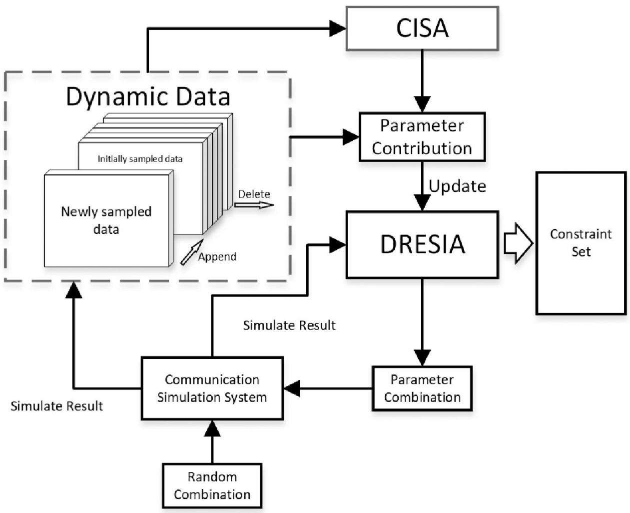


FIGURE 2. **Gray Box System.**

be coherent, which may leads to functional deviations and poor generalization. Hence, the FM-based White Box system in large-scale dynamic system cannot evolve in most cases.

In our work, we proposed a novel approach namely Gray Box to solve the shortcomings of the FM-based White Box system to achieve parameters' contribution analysis and constraint set inversion. Since the system will be affected by environmental changes, and the output will also change accordingly, which may cause the system output to fail to meet the requirements. Therefore, it is crucial to improve the evolvability of the model to ensure the stable output of the system. Thus our method focuses on the update of the measured data and how to search the constraint set. As shown in Fig. 2, CISA calculates the contribution of each parameter to the system output based on the existing data set. DRESIA trains the built-in agents with the existing data set and the parameter contribution set by DRL-based method. Then agent stop training when DRESIA always selects the parameter combinations with high system output. Thus we can obtain a series of parameter combinations when the output meets the system output requirement. To enable DRESIA to continuously output the parameter constraint set under the current operating environment, we let the agent continue to output parameter combinations and input them into the large-scale 5G wireless communication system to get new measured data and discard the same amount of origin data according to the First In First Out (FIFO) principle. Furthermore, to make the agent sample enough parameter combinations with high-output, and avoid the distortion of the output result of DRESIA caused by the agent sampling concentration, we additionally perform uniform sampling to supplement the data set. Finally, we update the Gray Box System with new dynamic data to realize the evolvability of the model. Table 2 shows a comparison between models.

TABLE I. **Notations**

Notations	Meanings
X	The system parameter set to be considered in this article
y	System output
y_{max}	The maximal system output
m	The length of the parameter set
n	The number of the sampled data
f_{ij}	Observations of parameters
μ	Power set
Z	The matrix of power sets
σ	The error variance of the linear multi-regression model
F_α	α-cut set
x_{span}	The value span of a single parameter
ω	Quantized value weight
ζ	List of differences in parameter values
Q	The quality function of Q-Learning
L	Loss function
θ	The network weights in DQN
s_t	The current state of agent
a_t	The current action of agent
r_t	The current reward that the Cyber-twin system feeds back to the agent
s_t	The next state of agent
A	Weight coefficient
p_{ij}	The change probability of i_{th} particle's j_{th} parameter value
Euc	Euclidean distance
c	The learning factor of PSO
X_{besti}	The individual historical best parameter combination of the i_{th} particle
X_{best}	The optimal parameter combination of the particle swarm
β	The inertia weight of PSO
AD	The inertia weight of PSO
γ	The discount factor

TABLE II. **Approach Comparison**

Feature / Model	Dynamic Data	Observability& Controllability	Model Evolvability
Black-Box	×	×	×
Fuzzy Measure	×	✓	×
Gray Box	✓	✓	✓

III. CHOQUET INTEGRAL-BASED INTERDEPENDENCY AND SIGNIFICANCE ANALYSIS ALGORITHM

Fuzzy measure is a powerful method to describe the cross-interactions, especially the non-linear Choquet integral. It has been successfully applied in evidence fusion and intelligence information processing [5], [9], [19], [26]. Since the massive MIMO system's throughput is a multi-dimensional factor influenced conception, the interaction among contribution of contributions towards the objective system throughput can be measured adequately through a non-additive FM, which is the main feature of the Choquet integral model.

Considering a set of attributes (system parameters in large-scale 5G simulation platform) $X = \{x_1, x_2, \ldots, x_m\}$, where

TABLE III. System Observation and Output

x_1	x_2	$\cdots$	x_m	y
f_{11}	f_{12}	$\cdots$	f_{1m}	y_1
f_{21}	f_{22}	$\cdots$	f_{2m}	y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
f_{n1}	f_{n2}	$\cdots$	f_{nm}	y_n

$m = |X|$. And n observed attributes together with their corresponding objective (system throughput) $y = \{y_1, y_2, \ldots, y_n\}$, then we have the form as Table 3.

The observation of attributes $x_1, x_2, \ldots, x_m$ can be regarded as a function $f : X \rightarrow (-\infty, +\infty)$, and the j_{th} observation of x_i is denoted by f_{ji}, where $1 \leq i \leq m$ and $1 \leq j \leq n$.

The interaction among attributes X towards the objective y is characterized a set function defined on the power set $\mu : P(X) \rightarrow \mathcal{R}$ with $\mu(\emptyset) = 0$, where $\mathcal{R}$ is the real domain. The Choquet integral-based model is defined as

$$y = \int_{(C)} f d\mu + \mathrm{N}(0, \sigma^2), \tag{1}$$

where $\mathrm{N}(0, \sigma^2)$ is a normally distribution random perturbation with expectation 0 and variance σ^2. Define an α-cut set F_α of f, where $F_\alpha = \{x | f(x) \geq \alpha\}$, for any $\alpha \in R$. The Choquet integral $\int_{(C)}$, with a fuzzy measure μ, is defined as

$$\int_{(C)} f d\mu = \int_{-\infty}^{0} [\mu(F_\alpha) - \mu(X)] d\alpha + \int_{0}^{+\infty} \mu(F_\alpha) d\alpha. \tag{2}$$

Obviously, when some parameters have a wide range of values, it will leads to a particularly disturbance that influence the Choquet integral's result. Thus, we should add a normalization method to preprocess the data in Table 3 with

$$x_{norm} = \frac{1 - e^{-\frac{2ex}{x_{span}}}}{1 + e^{-\frac{2ex}{x_{span}}}}, \tag{3}$$

where $x_{span} = |\max x_i - \min x_i|$.

A. SOLVING THE CHOQUET INTEGRAL WITH NON-ADDITIVE MEASURE

To solve the Choquet integral, according to [2], we need to reduce the non-linear multi-regression model to the linear multi-regression model by converting each m-dimensional vector attribute datum to a 2^m-dimensional vector datum, which is defined by Eq. (2). Then we have the linear multi-regression model

$$y = Z\mu + \mathrm{N}(0, \sigma^2), \tag{4}$$

where $Z = (z_{ij})_{n \times M}$, $M = 2^m - 1$, and $\mu = (\mu(\{x_1\}), \ldots, \mu(\{x_m\}), \mu(\{x_1, x_2\}) \cdots, \mu(\{x_1, x_2, \ldots, x_{m-1}\}), \mu(\{x_1, x_2, \ldots, x_m\}))^T$, $y = (y_1, y_2, \ldots, y_n)^T$.

In order to reduce the amount of calculation and storage space, [34] proposed a genetic algorithm to compute the Choquet integral, where Z is expressed as Algorithm 1.

Algorithm 1: Augmented Matrix Construct.

```
Input: The observation attributes f defined in Table 3
Outputs: The augmented matrix Z with size of n × (2^m + 1)
Initialize an empty matrix Z = [a_nk] with size of n × (2^m + 1), where k = 1, 2, ..., 2^m
for i = 0, n − 1 do
   Initialize Z[i, 0] = 0
   for j = 1, 2^m − 1 do
      if min_{j'=1}(f_ij) ≥ max_{j'=0}(f_ij) then
         Z[i, j] = min_{j'=1}(f_ij) − max_{j'=0}(f_ij) where j' ∈ [1, m]
      else
         Z[i, j] = 0
      end if
   end for
end for
```

After derived the matrix $Z \in R^{n \times M}$ for the determination of fuzzy measures, the matrix Z can be written as

$$\begin{aligned} Z &= (z_1^T, z_2^T, \ldots, z_n^T)^T \\ where \ \ z_i^T &= (z_{i1}, z_{i2}, \ldots, z_{iM}) \neq 0 \end{aligned} \tag{5}$$

Then we can solve the Eq. (4) by using the standard least-square method. For the given observation data, the optimal regression coefficients μ can be determined by using the least squares method in order to make σ^2 minimal. Therefore, the least square problem Eq. (4) can be written as the following equivalent form:

$$P_I = \min_{\mu} \|Z\mu - y\|, \tag{6}$$

where $\mu = (\mu_1, \mu_2, \ldots, \mu_M)^T$. According to [2], Wang *et al.* proposed a solution to Eq. (6):

$$\mu_{z_i}^T = z_i (z_i^T z_i)^{-1} y_i, \tag{7}$$

and update μ with

$$\mu_{i+1} = \mu_i + (1 - \gamma) \left[1 - \frac{\langle \mu, \mu_{z_i} \rangle}{\langle \mu_{z_i}, \mu_{z_i} \rangle} \right] \mu_{z_i}, \tag{8}$$

where $\gamma \in [0, 1]$.

With this procedure, one observation is processed with the least norm approach per time. Assume n observations is enough to approximate μ. This is reasonable because the number of rows in Z can be any number in the way of constructing matrix Z by randomly sampling the n observations. Thus the complete algorithm is shown as Algorithm 2.

In this algorithm, constructing the augmented matrix with Algorithm 1 takes the time of $O(nMm)$, where $M = 2^m - 1$. Calculating the solution with Algorithm 2 takes the time of $O(nMm)$. Therefore, the overall computation complexity is $O(nMm) + O(nM^2) = O(nM^2)$.

Algorithm 2: The Solution of the Least Squares Problem Eq. (6).

Require: The observations $\{(z_i, y_i)_{i=1}^n\}, \gamma, \epsilon$
$\mu_0 = 0$
while $i < n$ **do**
  Derive the solution of Eq. (6) with Eq. (7), i.e.,
    $\mu_{z_i} = z_i^T (z_i z_i^T)^{-1} y_i$
  Calculate the solution with
    $\mu_{i+1} = \mu_i + (1-\gamma)\left[1 - \frac{\langle \mu, \mu_{z_i}\rangle}{\langle \mu_{z_i}, \mu_{z_i}\rangle}\right]\mu_{z_i}$,
  if the last 100 μs do not change **then**
    Exit;
  end if
  i+=1
end while

B. SOLVING THE CHOQUET INTEGRAL WITH CISA APPROACH

For a system with a real time updating data set, its operating efficiency is sensitive to the computational complexity of a single step. The Eq. (8) will generate a lot of calculation and memory footprint. Thus we need a faster approach to solve the Choquet integral.

To solve Eq. (2) with lower complexity, we permute the indices so that the value of $\{f(x_1), f(x_2), \ldots, f(x_m)\}$ satisfies

$$f(x_{(0)}) \leq f(x_{(1)}) \leq f(x_{(2)}) \leq \cdots \leq f(x_{(m)}), \tag{9}$$

where $f(x_{(0)}) = 0$, and $x_{(1)}, x_{(2)}, \ldots, x_{(m)}$ is a permutation of $x_1, x_2, \ldots, x_m$.

Therefore, the discrete Choquet integral of a function $f : X \to R^+$ with respect to μ is

$$\int_{(C)} f d\mu = \sum_{i=1}^{m} \big(f(x_{(i)}) - f(x_{(i-1)})\big)\,\mu(\Theta_i), \quad where \;\; \Theta_i = \{x_{(i)}, x_{(i+1)}, \ldots, x_{(m)}\}, \tag{10}$$

The specific steps to solve the discrete Choquet integral is presented in Appendix A. According to Appendix A, for observed attribution z_i^T, there are at most m additive actions on the system throughput formation with Eq. (7), then Eq. (7) can be rewritten as

$$\mu_\zeta^T = \zeta_i(\zeta_i^T \zeta_i)^{-1} y_i, \quad where \;\; \zeta_i^T = \{w_1(f_i(x_{(1)}) - f_i(x_{(0)})), w_2(f_i(x_{(2)}) - f_i(x_{(1)})), \ldots, w_m(f_i(x_{(m)}) - f_i(x_{(m-1)}))\}, \tag{11}$$

and w_j is an adjustable weight parameter. When $w_j = 1$, the result of the Gray Box model reflects the real impact of parameter combinations to the system output. When w_j increases with the result of $f_i(x_{(j)}) - f_i(x_{(j-1)})$ increases, the parameter combination that has a large impact on the system output will have a higher contribution value. By adjusting the value of w_j, we can control the accuracy and universality of the constraint set. The specific analysis is introduced in Section.V.

Now we can obtain the parameter combination that has the greatest impact on the system throughput by solving the Choquet Integral with Algorithm 3.

Algorithm 3: Choquet Integral-Based Interdependency and Significance Analysis Algorithm.

Input: Pre-processed ζ and objective y, hyperparam ϵ
Outputs: The solution μ that represents each parameter combination's contribution to the output
Initialize an index list L to storage origin index of each parameter
for $i = 1, n$ **do**
  Sort the i_{th} observed attribution and storage the sorted index and corresponding data L_{sort} and f_{sort}
end for
for $i = 1, n$ **do**
  Derive the solution with $\mu_\zeta^T = \zeta_i(\zeta_i^T \zeta_i)^{-1} y_i$
  for j=1,m **do**
    Compare the L and L_{sort} to find the $\zeta_{i,j}$'s corresponding subset and update is with $\mu = \epsilon\mu + (1-\epsilon)\mu_\zeta$
  end for
end for

For solving the Choquet integral with Eq. (11), the computation complexity of deriving $\mu(z_i)$ will be reduced from $O(M)$ to $O(m)$. So the overall complexity is reduced from $O(nM)$ to $O(nm)$, where $M = 2^m - 1$.

IV. DEEP REINFORCEMENT LEARNING-ENABLED CONSTRAINT SET INVERSION ALGORITHM

Q-Learning (QL) is a value-based algorithm [29], and it is one of the most important breakthroughs in reinforcement learning. Since QL is a tabular method, in the case of complex states and actions, the Q table in QL will take exponential time to converge, thus the applicable state and action space explored in a given period time is very small. On the other hand, if a state never appears, QL cannot handle it, which means it has low generalization ability.

To break through the QL's bottleneck, Deep Q-Learning uses a Deep Q-Network (DQN) to fit Q_table:

$$Q(s, a; \theta) \approx Q^*(s, a). \tag{12}$$

The optimization goal of the DQN is to minimize the square loss of 1-step TD error

$$L = \mathbb{E}\left[(r + \gamma \max_{a'} Q(s', a') - Q(s, a))^2\right], \tag{13}$$

leading to the QL gradient

$$\frac{\partial L(\theta)}{\partial \theta} = \mathbb{E}(r + \gamma \max_{a'} Q(s', a') - Q(s, a))\frac{\partial Q(s, a, \theta)}{\partial \theta}, \tag{14}$$

thus the DQN can be optimized by Stochastic Gradient Descent (SGD).

After obtaining the contribution of the parameter combination to the system output, for parameter combinations with high contributions, we can easily adjust the value of the parameter to improve the system output. However, the reasons for the parameters that have little contribution to the experimental results are not single. For instance, insufficient and unevenly distributed sampled data may result in some parameter combinations having no corresponding valid results. Besides, in the case of the control variate method, some parameters will not show monotonicity as the value increases or decreases. These scenarios make the process of constraint set inversion very difficult.

To handle these situations, we adopt DNN to fit the corresponding influence of unsampled parameter combinations on the system output. Hence, the DQN can solve the problem of discontinuity in the state space composed of parameter configuration, so as to quickly search for the feasible region of the system.

1) THE DQN FORMULATION

State Space: Suppose there has n parameters $\{V_1, V_2, \ldots, V_n\}$ that affect system output, each parameter combination $\{V_{(1,randn)}, V_{(2,randn)}, \ldots, V_{(n,randn)}\}$ is defined as a state s, where the 'randn' is a uniform random sampling in each parameter's ranges.

Action Space: To simplify the MDP state transition process, the value change of each parameter is represented by value increase or value decrease in each iteration, thus we define $a_t = 1$ as value increase and $a_t = 0$ as value decrease. Then we have

$$\begin{aligned} s_{t+1} &\leftarrow \{s_t, a_t\} \\ &= \{(V_{(1,randn)}, a_{t1}), (V_{(2,randn)}, a_{t2}), \\ &\qquad (V_{(3,randn)}, a_{t3}), \ldots, (V_{(n,randn)}, a_{tn})\}. \end{aligned} \tag{15}$$

The action is determined by the Q value fitting by DNN:

$$a_t = \arg\max_a Q_\theta(s_t, a) \tag{16}$$

Reward: In order to solve the problem of non-dense data sets and high system simulation cost, the reward function is calculated by system output y and corresponding value of contribution μ_y obtained by Algorithm 3. To balance the efficiency and accuracy of the feasible region, we find the parameter combination in the measured dataset that has the minimum Euclidian distance to the parameter combination represented by the current State in 75% of the time, then calculated the reward by combing its corresponding system output and its result of Choquet integral. In the remaining 25% of the time, we input the parameter combination represented by current State into Cyber-twin system simulation platform to get the real system output, and then calculate the corresponding reward.

$$r_t = \frac{Ay}{y_{max}} \cdot \mu_y, \tag{17}$$

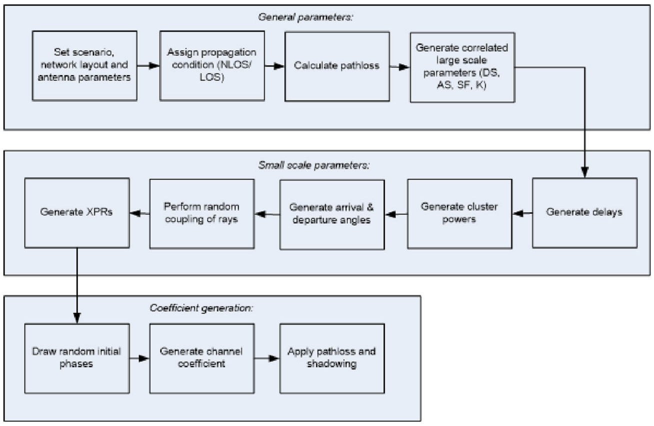


FIGURE 3. Modeling process for the Cyber-twin system.

where A is a weight coefficient set according to the actual situation. If s_{t+1} does not exist in the measured data set, we use the state which has the minimum Euclidean distance E_d and its output to replace the y. If the minimum E_d is still too large to influence the result, then set the $r_t = 0$, the threshold depends on the actual situation. When this situation happens, input s_{t+1} into the simulation system, and use the result to update the data set.

1) TRAINING

Based on the above formulation, we using a DNN to represent the $Q_\theta(s_t, a)$ as Q-network which takes as input the state s_t, and outputs a set of Q value to represent the value of each element in s_t after taken different actions. The Q-network is trained with a variant REINFORCE algorithm in an episodic setting. In the training session, we run M episodes for a fixed duration of T iterations to explore the various action's reward using the current DNN's output Q value. The transition consisted of s_t, a_t, r_t, s_{t+1} in every iteration will store in the memory D. Once the memory D is full, sample a minibatch from D to update the weights of DNN with the gradient descent method. The agent continuously explores the rewards corresponding to different state s_{t+1}, until the output corresponding to the explored s_{t+1} meets the requirements and reward converged at the same time, then performs the next episode training. The implementation of the variant REINFORCE algorithm is described as Algorithm 4.

V. EXPERIMENT

A. CYBER-TWIN SIMULATION PLATFORM

The Cyber-twin system is based on the 3D channel model standard specified in 3GPP TR 38.901 protocol [13]. The specific modeling process can be seen in the Fig. 3, all kinds of large scale and small scale parameters are considered and included. The channel model used in this article has been corrected, and the benchmark data is from 3GPP R1-140843. The data and average values of 9 major companies in the industry are compared, and the CDF curve of the platform output results is basically consistent with the average data of

Algorithm 4: Deep Reinforcement Learning-Enabled Constraint Set Inversion Algorithm.

Input: The hyper-parameter A and Threshold
Initialize memory D with the size of N;
Initialize the evaluation network Q_{eval}, the target network Q_{target} with random weights;
for episode=1, M **do**
  Initialize the state $\{V_{(1,randn)}, \ldots, V_{(n,randn)}\}$ randomly from the state space;
  Initialize the counter=0;
  while True **do**
    With the probability ϵ select an action $a_t = \arg\max_a Q_\theta(s_t, a)$, otherwise select a random action;
    Fetch the $s_{t+1} \leftarrow \{s_t, a_t\}$ and find most similar state in data set;
    if E_d <=Threshold **then**
      calculate the reward $r_t = \frac{Ay}{y_{max}} \cdot \mu_y$;
    else
      set reward $r_t = 0$, and input state s_{t+1} to simulation system to update the data set;
    end if
    Store or cover transition (s_t, a_t, r_t, s_{t+1}) in memory D;
    if D is full **then**
      Sample random minibatch of transitions (s_t, a_t, r_t, s_{t+1}) from memory D;
      Train the DNN with gradient descent $\frac{\partial L(\theta)}{\partial\theta} = \mathbb{E}(r + \gamma \max_{a'} Q(s', a') - Q(s, a))\frac{\partial Q(s,a,\theta)}{\partial\theta}$;
    end if
    if r_t converged or counter > T **then**
      Break;
    end if
    counter+=1;
    $s_t = s_{t+1}$;
  end while
end for

companies. The interference scheme adopts an additive noise model, and the interference power of other users sharing the resource block is calculated based on the accumulation of the resource block scheduled by the user. The interference power is calculated according to the channel fading coefficient of the interference user corresponding to the TTI calculated by the 3D channel model and the precoding matrix. The minimum granularity of the resource block is at the sub-carrier level. The white Gaussian noise model was used for the noise.

B. SIMULATION SETUP

The number of antennas at the base station (BS) is 16, and its arrangement is 8×2 planar array. We use the JSDM [39] scheme as the precoding mode, and use standard defined channel quality indication (CQI) as a link mapping scheme.

TABLE IV. Parameters for Choquet Integral-Based Algorithm

Parameter	Value
Gray Box weight parameter w	[1, 2]
Antenna Configuration at BS	Number:16, 8×2 area array
Maximum UT Number	10
Precoding Mode	JSDM
Link Mapping Scheme	CQI
Downlink Channel quantitative feedback scheme	Correlation statistical model
System Bandwidth (MHz)	$\{1, 2, 3, \ldots, 19, 20\}$
BS Antenna Array Downtilt (Degree)	$\{0, 1, 2, \ldots, 29, 30\}$
BS Antenna Array Vertical Spacing (Number of wavelength, here one wavelength is equal to 15cm)	$\{0.6, 0.7, 0.8, 0.9, 1.0, 1.1\}$
Maximal Scheduled UT under MU-MIMO System Mode	$\{1, 2, 3, 4\}$
Maximum and Minimum Ratio Threshold of Channel Matrix Eigenvalues	$\{10, 11, 12, \ldots, 999, 1000\}$
Massive-MIMO Sample Span (Number of subcarriers)	$\{1, 2, 3, 4, 6, 12\}$
Channel Feedback Delay of JSDM Algorithm of MU-MIMO (millisecond)	$\{0, 1, 2, 3, 4\}$
UT Speed (km/h)	$\{0, 1, 2, \ldots, 89, 90\}$

TABLE V. Parameters for DRESIA

Parameter	Value
State Size	8
Action Size	Number:16, Form:8×2
Discount Factor γ	0.9
DNN Learning Rate l_r	0.01
Action Probability ϵ	0.85
Memory D Capacity	2000
Batch Size	32
Target Network Update Iteration	100
DNN's Hidden Layer Structure	$\{32, 128, 256, 128, 32\}$

Adopt statistical correlation model as the quantitative feedback scheme of the downlink channel. To verify the influence of other antenna settings on the system throughput, other parameters needed for the simulation are listed in Table 4.

Once the simulation parameters are set, the value will be randomly sampled to form a series of parameter combinations X. After obtaining the output y after system simulation, we can get the contribution degree μ of the parameter combination to the output through Algorithm 3.

For DRESIA, the size of the state space is the same as the size of the parameter combination X, which is 8. The size of the action space is 16 with the form of 8×2, corresponding to the increase or decrease of each state in the state space. The discount factor γ is set to 0.9. For DNN's structure, the size of the input layer and the output layer is the same as the size of the state space and the size of the action space. And we use 5 hidden layers consisted {32, 128, 256, 128, 32} neurons to learn and approximate the quality value. Other parameters needed for the simulation are listed in Table 5.

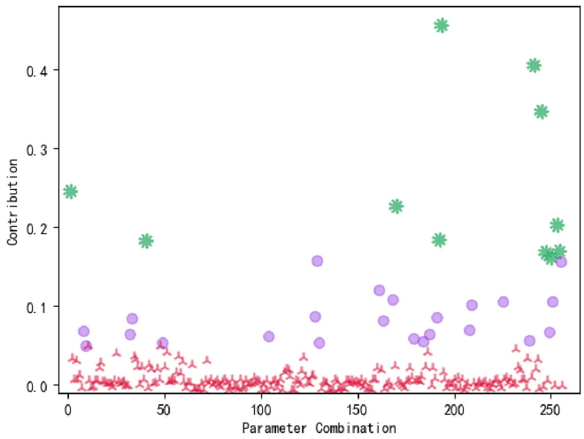


FIGURE 4. Non-additive measure vector μ.

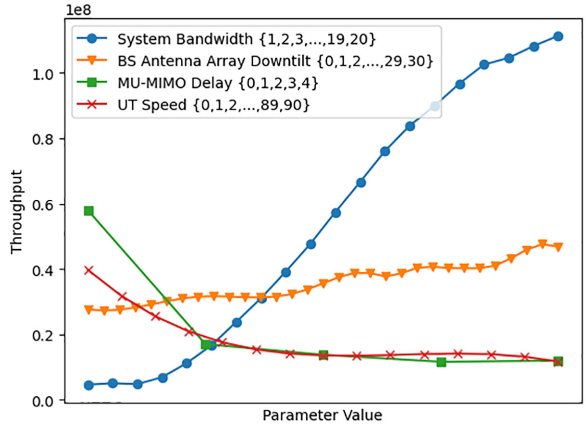


FIGURE 5. The relationship between throughput and parameter value.

C. METHOD VERIFICATION

We will verify the feasibility of the two algorithms separately in this subsection.

1) CHOQUET INTEGRAL-BASED INTERDEPENDENCY AND SIGNIFICANCE ANALYSIS ALGORITHM

We randomly sampled these 8 parameters X as mentioned in Table 4 for n times and input them into the large-scale 5G simulation platform to get the corresponding system average throughput y as sampled paths. Since each path contains 8 option points, we derive the coefficient matrix Z with Algorithm 1. Thus, we have a least square problem as

$$\min_{\mu} \|Z\mu - y\|, \tag{18}$$

where Z is the $n \times 256$ matrix, μ is the 256×1 vector, and y is the $n \times 1$ vector. Then we implement the Algorithm 3 to derive the non-additive measure vector μ as shown in Fig. 4. The parameter combinations marked in green will be multiplied by the maximum weight w, and the parameter combinations marked in purple and red will have smaller weights.

As mentioned in the above analysis, the higher the value of the contribution, the greater the impact on the system. And we found the $195th$ element of the vector μ has the highest contribution, which represents the parameter combination $\{x_1, x_2, x_7, x_8\}$, and its binary form is $\{1, 1, 0, 0, 0, 0, 1, 1\}$, corresponding to System Bandwidth, BS Antenna Array Downtilt, Channel Feedback Delay of JSDM Algorithm of MU-MIMO and UT Speed these 4 parameters. To verify the accuracy of the results, we use the control parameter method to verify the influence of these parameters on the system throughput. The result is shown in Fig. 5.

Refer to Fig. 5, we can find that the system throughput is basically monotonic with the changes of these parameters.

Thus, we only adjusting these selected parameters. The options for the other parameters are randomly selected from the related ranges. By only adjusting these 4 in 8 parameters, the system throughput is 108,841,764 bit/s, 90.295% of the maximum throughout (120,540,409 bit/s) by adjusting all 8 parameters.

2) DRESIA

In the previous experiment, CISA can extract the contribution of each parameter to the system QoS through a small amount of uniformly sampled measured data, and great QoS can be obtained by simply adjust the system parameters based on these data. However, in consideration of generalization, deploying such a parameter adjustment scheme in a real scenario does not have strong robustness. Therefore, we need to attach an inversion algorithm to determine how the parameters are configured.

To ensure that the data used for testing is highly reliable, the large-scale 5G simulation system we adopted simulates the multi-user and multi-cell communication system and the internal working mechanism of the system through a system-level simulation method, which can realize the evaluation of the key performance indicators of the system. It calculates the channel fading characteristics based on the ray tracing data, and calculates the signal to interference noise ratio and throughput based on the channel fading characteristics. Multi-core parallel simulation and hardware acceleration simulation technology are used to realize an efficient 5G system simulation platform, calculate the channel matrix and precoding matrix, and further realize the evaluation of performance such as interference and system capacity. After deploying Algorithm 4 to train the agent, we use the trained model to explore the parameter configuration that meets the output requirements. First, we randomly select three different parameter combinations as the initial state s_1. Then we input the state s_1 into the trained model with the probability of ϵ to get the action a_1 with the highest Q value, or randomly choose an action as a_1. As shown in Fig. 6, after a period of iteration, the reward $\{r_1, r_2, \ldots, r_t\}$ shows an upward trend. Due to the difference in the initial state, the initial reward is also different, but in the end, they both converged.

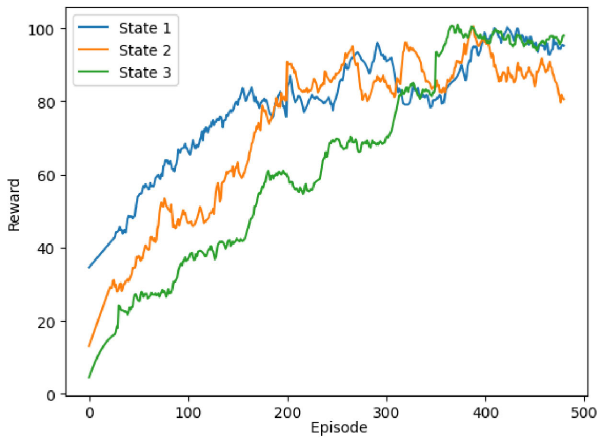


FIGURE 6. **Reward of 3 initial states (ϵ=0.85).**

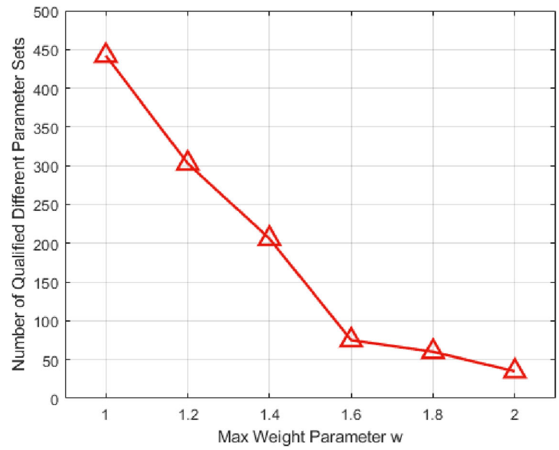
TABLE VI. **Correctness of Agent's Selection Under Different Requirements**

Baseline	100%	95%	90%	85%	80%
Accuracy	6.05%	15.39%	44.72%	84.66%	97.20%

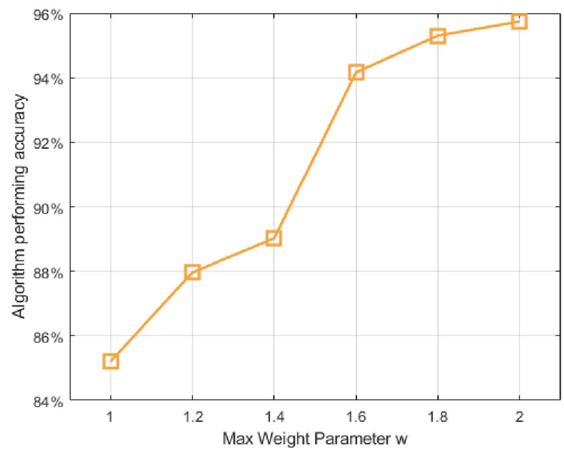
To verify the effectiveness of the algorithm more intuitively, we sample the state corresponding to the converged reward at equal intervals and input it to the large-scale 5G simulation platform, and verify whether the state s selected by the agent meets the standard according to the output throughput rate.

Five baselines that represent 80%∼100% of the maximum throughput (120,540,409 bit/s) are set to measure the quality of the results selected by the agent. We input the parameter configuration selected by the agent into the large-scale 5G simulation platform, then count the number of parameter configurations that meet different standards and records in Table 6. We can observe from Table 6, when the standard is 85% of the optimal system throughput (about 100,000,000 bit/s), DRL-enabled constraint set inversion algorithm can output the parameter combinations that meet the requirements with the probability of 84.66%. Moreover, when the standard is 80% of the optimal system throughput (about 96,000,000 bit/s), DRL-enabled constraint set inversion algorithm can output the parameter combinations that meet the requirements with the probability of 97.20%.

The Black Box-based method is scenario-dependency, and training a brand new model for each time node is time-consuming. Therefore, several White Box-based heuristic algorithms, including GA [22], SA [8] and ACO [31] are compared and discussed. An improved PSO [24] that combines the advantages of both PSO and ACO is also compared and discussed, the detailed algorithm refers to Appendix B. The experimental results in static environment are shown in Table 7. For operation efficiency, compared with the Gray Box method. The White Box-based method takes more time to perform each episode, especially the ACO algorithm takes over 100 s. The trained Gray Box model can quickly output the parameter configuration in the current



(a)



(b)

FIGURE 7. **The relationship between weight and the correct rate and quantity of parameter combinations. (a) The relationship between the number of qualified different parameter combinations and the weight parameter w. (b) The relationship between the accuracy of output parameter combinations and the weight parameter w.**

environment according to the input. More efficient algorithms can achieve algorithm adaptation faster. The phenomenon of the White Box-based method output repeated parameter combinations that meets the system output requirements is very common. The SA algorithm has the most different effective parameter combinations but the lowest accuracy, whereas the ACO has the opposite situation. In comparison, the Gray Box method has the best performance with 442 different effective parameter combinations in 500 iterations. Since the GA, SA, ACO and the improved PSO algorithms are entirely the White Box-based algorithms, they have higher accuracy than DRESIA when the weight parameter $w = 1.0$ as mentioned in Section III due to the real-time data feedback.

Considering that under actual conditions, the probability of congestion in the wireless communication network at midnight is small, and the change in service density, bandwidth requirements is also small. In this case, the compliance rate of

TABLE VII. Method Efficiency and Accuracy Comparison

Index \ Method	White Box-based Methods				Gray Box Method	
	GA	SA	ACO	Improved PSO	DRESIA(w=1)	DRESIA(w=2)
Running Time for 500 Iterations	≈308min	≈305min	≈1433min	≈458min	≈3min	
Compliance Rate in 500 Iterations	96.0%	95.4%	97.4%	96.8%	89.2%	95.8%
Number of Qualified Diff Param Sets in 500 ITERs	21	49	13	16	442	35

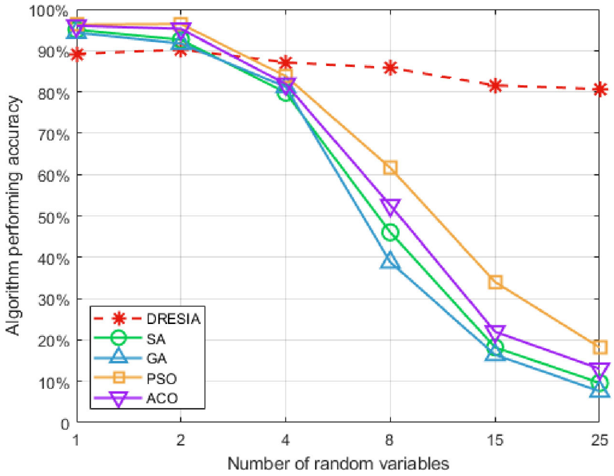


FIGURE 8. The accuracy in dynamic large-scale 5G simulation platform.

the output parameter configuration of DRESIA is not as high as that of the White-Box method, but we want to retain the advantage of fast parameter configuration output of DRESIA. By changing the weight parameter w, we can therefore adjust the output of DRESIA to favor high robustness in a dynamic environment or high accuracy in a static environment.

With the max weight parameter w increases, the agent will have a higher probability to choose the optimal solution. Thus, the number of the available parameter configurations explored becomes less. As shown in Table 7 and Fig. 7, when w becomes larger, the compliance rate of output parameter configuration approaches the White Box-based method, which proves that DRESIA has good generalization.

To meet the actual dynamic environment, we set some parameters of the large-scale 5G simulation platform fluctuate in a small range, and run the White Box-based algorithms until convergence and then iterate 500 times, respectively. The accuracy of Gray Box can be directly expressed using the results of the above experiment since the experiment was done in the dynamic environment. As shown in Fig. 8, we can find that the accuracy of the White Box-based algorithms immediately declines (7.60%, 9.60%, 13.4% and 18.2% with 25 randomly configured environmental parameters, respectively) because the global optimal solution has changed due to the dynamics of the system. Since the data set of Gray Box is continually updated, and the Q-network is continually being updated, the accuracy of Gray Box has not changed significantly (84.66% with 25 randomly configured environmental parameters), which means it has the strong adaptive ability and is therefore more suitable for deployment in the real environment. Furthermore, Gray Box has a faster feasible region search speed. Hence, the Gray Box method has absolute advantages in dynamic system.

VI. CONCLUSION

In this article, a novel Gray Box-based approach namely DRESIA that contains CISA algorithm has been proposed and analyzed. Different from traditional methods based on White-Box or Black-Box that facing the problem of high complexity and low generalization, our new method does not require massive measured data in different time periods, and can obtain the constraint set in dynamic environment by sampling real-time data with low frequency. We compared multiple criteria for DRESIA, SA, GA, PSO and ACO, which include the accuracy of the feasible region, the time for the algorithm to output a single parameter combination, the number of different parameter combinations satisfying the given output index, and the accuracy of the output results under dynamic environment as indicators. The results show that the proposed Gray Box-based approach DRESIA is more stable in the dynamic environment compared with other algorithms. The complexity of DRESIA and the required data are much lower than White box-based methods. Meanwhile, and it has better generalization ability than the Black box-based and White box-based methods. Therefore, DRESIA is suitable for large-scale dynamic system.

REFERENCES

[1] A. M. Ahmed, A. H. A. Hashim, and W. H. Hassan, "Investigation of gateway placement optimization approaches in wireless mesh networks using genetic algorithms," in *Proc. Int. Conf. Comput. Commun. Eng.*, 2014, pp. 68–71.

[2] W. An, C. Ren, S. Ci, D. Wu, H. Luo, and Y. Liu, "A novel low-complexity method for determining nonadditive interaction measures based on least-norm learning," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2014, pp. 1705–1712.

[3] A. Anurag, R. Priyadarshi, A. Goel, and B. Gupta, "2-D coverage optimization in wsn using a novel variant of particle swarm optimisation," in *Proc. 7th Int. Conf. Signal Process. Integr. Netw.*, 2020, pp. 663–668.

[4] J. A. Ayala-Romero, J. J. Alcaraz, J. Vales-Alonso, and E. Egea-López, "Online optimization of interference coordination parameters in small cell networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6635–6647, Oct. 2017.

[5] L. Baccour and A. M. Alimi, "Distance measures for intuitionistic fuzzy sets and interval valued intuitionistic fuzzy sets," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2019, pp. 1–6.

[6] S. Ci and H. Guo, "Quantitative dynamic interdependency measure and significance analysis for cross-layer design under uncertainty," in *Proc. 16th Int. Conf. Comput. Commun. Netw.*, 2007, pp. 900–904.

[7] C. Crespi, R. A. Scollo, and M. Pavone, "Effects of different dynamics in an ant colony optimization algorithm," in *Proc. 7th Int. Conf. Soft Comput. Mach. Intell.*, 2020, pp. 8–11.

[8] H. Daryanavard and A. Harifi, "Uav path planning for data gathering of iot nodes: Ant colony or simulated annealing optimization," in *Proc. 3rd Int. Conf. Internet Things Appl.*, 2019, pp. 1–4.

[9] S. Das, D. Guha, and R. Mesiar, "Information measures in the intuitionistic fuzzy framework and their relationships," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 3, pp. 1626–1637, Jun. 2018.

[10] S. Dastoor, U. Dalal, and J. Sarvaiya, "Comparative analysis of optimization techniques for optimizing the radio network parameters of next generation wireless mobile communication," in *Proc. 14th Int. Conf. Wireless Opt. Commun. Netw.*, 2017, pp. 1–6.

[11] D. Devaraj and B. Yegnanarayana, "Genetic-algorithm-based optimal power flow for security enhancement," *IEE Proc.-Gener. Transmiss. Distribution*, vol. 152, no. 6, pp. 899–905, Nov. 2005.

[12] M. Eckhart, A. Ekelhart, and E. Weippl, "Enhancing cyber situational awareness for cyber-physical systems through digital twins," in *Proc. 24th IEEE Int. Conf. Emerg. Technol. Factory Automat.*, 2019, pp. 1222–1225.

[13] K. Flynn, "Release 15. internet," [Online]. Available: https://www.3gpp.org/release-15.

[14] J. Han, D. Choi, S. Park, and S. Hong, "Hyperparameter optimization for multi-layer data input using genetic algorithm," in *Proc. IEEE 7th Int. Conf. Ind. Eng. Appl.*, 2020, pp. 701–704.

[15] T. C. Havens and D. T. Anderson, "Machine learning of choquet integral regression with respect to a bounded capacity (or non-monotonic fuzzy measure)," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2019, pp. 1–6.

[16] T. C. Havens, D. T. Anderson, and C. Wagner, "Data-informed fuzzy measures for fuzzy integration of intervals and fuzzy numbers," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 5, pp. 1861–1875, Oct. 2015.

[17] A. A. Kannan, G. Mao, and B. Vucetic, "Simulated annealing based localization in wireless sensor network," in *Proc. IEEE Conf. Local Comput. Netw. 30th Anniversary*, 2005, pp. 2 pp.–514,.

[18] M. M. Keikha, "Improved simulated annealing using momentum terms," in *Proc. 2nd Int. Conf. Intell. Syst., Modelling Simul.*, 2011, pp. 44–48.

[19] Z. Li, P. Zhang, X. Ge, N. Xie, G. Zhang, and C. Wen, "Uncertainty measurement for a fuzzy relation information system," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 12, pp. 2338–2352, Dec. 2019.

[20] C. Liang and Y. Lin, "A coverage optimization strategy for mobile wireless sensor networks based on genetic algorithm," in *Proc. IEEE Int. Conf. Appl. Syst. Invention*, 2018, pp. 1272–1275.

[21] Y. Liu, W. Huangfu, H. Zhang, and K. Long, "Elite gradient descent optimization of antenna parameters constrained by radio coverage in green cellular networks," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2018, pp. 843–847.

[22] B. Lorenzo and S. Glisic, "Optimal routing and traffic scheduling for multihop cellular networks using genetic algorithm," *IEEE Trans. Mobile Comput.*, vol. 12, no. 11, pp. 2274–2288, Nov. 2013.

[23] G. Molina and E. Alba, "Wireless sensor network deployment using a memetic simulated annealing," in *Proc. Int. Symp. Appl. Internet*, 2008, pp. 237–240.

[24] W. Peng, Z. Yang, C. Liu, J. Xiu, and Z. Zhang, "An improved pso algorithm for battery parameters identification optimization based on thevenin battery model," in *Proc. 5th IEEE Int. Conf. Cloud Comput. Intell. Syst.*, 2018, pp. 295–298.

[25] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21980–22012, 2020.

[26] V. Singh, S. P. Yadav, and R. Mesiar, "Information measures in atanassov's intuitionistic fuzzy environment and their application in decision making," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 11, pp. 2905–2917, Nov. 2020.

[27] M. Soleimanpour-Moghadam and H. Nezamabadi-Pour, "Discrete genetic algorithm for solving task allocation of multi-robot systems," in *Proc. 4th Conf. Swarm Intell. Evol. Comput.*, 2020, pp. 00 6–009.

[28] X. Sun, Z. Shi, and J. Zhu, "Multi-objective design optimization of an IPMSM for EVs based on fuzzy method and sequential taguchi method," *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, Nov. 2021, pp. 10592–10600, doi: 10.1109/TIE.2020.3031534.

[29] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A. Bradford Book, 2018.

[30] S. Tesic, A. Cica, B. Sredanovic, and B. Knezevic, "Application of taguchi method for identifying optimum specific energy consumption in cnc face milling," in *Proc. 19th Int. Symp. INFOTEH-JAHORINA*, 2020, pp. 1–4.

[31] L. V. Tung, L. H. Manh, C. D. Ngoc, M. Beccaria, and P. Pirinoli, "Automated design of microstrip patch antenna using ant colony optimization," in *Proc. Int. Conf. Electromagnetics Adv. Appl.*, 2019, pp. 0587–0590.

[32] S. Wang and C. Ran, "Rethinking cellular network planning and optimization," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 118–125, Apr. 2016.

[33] Z. Wang and R. M. Jungers, "A data-driven method for computing polyhedral invariant sets of black-box switched linear systems," *IEEE Contr. Syst. Lett.*,, vol. 5, no. 5, pp. 1843–1848, Nov. 2021.

[34] Z. Wanga and H. Guo, "A new genetic algorithm for nonlinear multiregressions based on generalized Choquet integrals," in *Proc. 12th IEEE Int. Conf. Fuzzy Syst., Vols 1, 2, IEEE Int. Conf. Fuzzy Syst.*, O. Nasaoui, H. Frigui, and J. M. Keller, Eds. IEEE; IEEE Neural Networks Soc, 2003, pp. 819–821.

[35] J.-T. Tsai, T.-K. Liu and J.-H. Chou, "Hybrid Taguchi-genetic algorithm for global numerical optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 4, pp. 365–377, Aug. 2004, doi: 10.1109/TEVC.2004.826895.

[36] D. Wu, S. Ci, H. Luo, and H.-F. Guo, "A theoretical framework for interaction measure and sensitivity analysis in cross-layer design," *ACM Trans. Model. Comput. Simul.*, vol. 21, no. 1, pp. 1–26, Dec. 2010.

[37] R. Yang, Z. Wang, P. Heng, and K. Leung, "Classification of heterogeneous fuzzy data by choquet integral with fuzzy-valued integrand," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 5, pp. 931–942, Oct. 2007.

[38] Y. Yang *et al.*, "Parallel channel sounder for MIMO channel measurements," *IEEE Wireless Commun.*, vol. 25, no. 5, pp. 16–22, Oct. 2018.

[39] Y. Yang *et al.*, "Decco: Deep-learning enabled coverage and capacity optimization for massive MIMO systems," *IEEE Access*, vol. 6, pp. 23361–23371, 2018.

[40] Y. Yang, Y. Li, W. Zhang, F. Qin, P. Zhu, and C. Wang, "Generative-adversarial-network-based wireless channel modeling: Challenges and opportunities," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 22–27, Mar. 2019.

[41] Z. M. L. J. Yu Wenxin, L. Kai and Y. Yang, "Research on adaptive neural network for 3D channel amplitude prediction based on ray-tracing data," *J. Univ. Chin. Acad. Sci.*, 2020.

[42] X. Zhang, X. Zhang, and Z. Wu, "Utility- and fairness-based spectrum allocation of cellular networks by an adaptive particle swarm optimization algorithm," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 4, no. 1, pp. 42–50, Feb. 2020.

[43] Q. Zhao and C. Li, "Two-stage multi-swarm particle swarm optimizer for unconstrained and constrained global optimization," *IEEE Access*, vol. 8, pp. 124905–124927, 2020.

[44] F. Zhu, Y. Lv, Y. Chen, X. Wang, G. Xiong, and F. Y. Wang, "Parallel transportation systems: Toward IoT-enabled smart urban traffic control and management," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4063–4071, Oct. 2020.

[45] X. Zhu and Y. Zhang, "Wireless sensor network path optimization based on particle swarm algorithm," in *Proc. IEEE Int. Conf. Comput. Sci. Automat. Eng.*, vol. 3, 2011, pp. 534–537.

[46] P. C. C. Zuniga, "An ant colony optimization for the network inference and parameter estimation of s-systems," in *Proc. 3rd Int. Conf. Biomed. Eng. Inform.*, vol. 6, 2010, pp. 2363–2367.

ZHOUYANG LIN received the B.Eng. degree in communication engineering from Xi'dian university, Xi'an, China, in 2019. He is currently working toward the M.A.Eng. degree with ShanghaiTech University, School of Information Science and Technology, Shanghai, China.

His research interests include artificial intelligence in wireless communication domain, financial big data, and the resource management and scheduling in fog networks.

KAI LI received the M.Eng. degree in signal and information processing from Tongji University, Shanghai, China, in 2000. He is currently a Senior Engineer of ShanghaiTech University. He is in charge of designing and developing a new generation of wireless simulation and edge computing platform. In the past 6 years, his team continued to build 5G simulation platform and submitted more than 30 papers, patent applications and standard proposals. He has more than 15 years of communication industry design and development experience. His research interests focus on large-scale communication system simulation and digital twin, edge computing, and AI enabled network planning and optimization technology. He has completed a number of successful research cooperation projects with Huawei, China Telecom, Sony and other enterprises in recent 8 years.

YANG YANG (Fellow, IEEE) received the B.S. and M.S. degrees in radio engineering from Southeast University, Nanjing, China, in 1996 and 1999, respectively, and the Ph.D. degree in information engineering from the Chinese University of Hong Kong in 2002.

He is currently a Full Professor with School of Information Science and Technology, Master of Kedao College, and Director of Shanghai Institute of Fog Computing Technology (SHIFT), ShanghaiTech University, China. He is also an Adjunct Professor with the Research Center for Network Communication, Peng Cheng Laboratory, China, as well as a Senior Consultant for Shenzhen SmartCity Technology Development Group, China. Before joining ShanghaiTech University, he has held faculty positions with the Chinese University of Hong Kong, Brunel University, U.K., University College London (UCL), U.K., and SIMIT, CAS, China. Yang's research interests include fog computing networks, service-oriented collaborative intelligence, wireless sensor networks, IoT applications, and advanced testbeds and experiments. He has published more than 300 papers and filed more than 80 technical patents in these research areas. He has been the Chair of the Steering Committee of Asia-Pacific Conference on Communications (APCC) since January 2019. In addition, he is a General Co-Chair of the IEEE DSP 2018 conference and a TPC Vice-Chair of the IEEE ICC 2019 conference.

FANGLEI SUN received the B.Eng. and M.Eng. degrees in computer science from Xi'dian University, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree in electronic and electrical engineering from the University of Hong Kong, Hong Kong.

She is currently a Research Associate Professor with ShanghaiTech University, Shanghai, China. Before joining ShanghaiTech University, she was a Research Scientist with Bell Labs, Nokia Shanghai Bell, Shanghai, China. She has authored or coauthored 20 papers, and filed more than 50 patents in wireless communications. Her current research interests include 5G and B5G networks, intelligent design for communication systems, and intelligent multiagent collaboration technologies.

LIANTAO WU received the B.Eng. degree in automation from Shandong University, China, in 2012 and the Ph.D. degree in control science and engineering from Zhejiang University, China, in 2017.

He is currently a Research Assistant Professor with the Shanghai Institute of Fog Computing Technology, ShanghaiTech Univesity, Shanghai, China. His research interests include IoT, wireless communications, and fog computing.

PANPAN SHI received the B.E. degree in communication engineering from Xi'dian university, Xi'an, China, in 2020. She is currently working toward the M.A.Eng. degree with ShanghaiTech University, School of Information Science and Technology, Shanghai, China.

Her research interests include edge computing, artificial intelligence in wireless communication domain.

SONG CI (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Nebraska-Lincoln, Lincoln, NE, USA, in 2002.

He is currently a Professor with Electrical Engineering Department, Tsinghua University, Beijing, China. Before joining Tsinghua University, he was a Tenured Associate Professor with the ECE Department, University of Nebraska-Lincoln. From 1992 to 1995, he was an Assistant Engineer with Shandong P&T Engineering Company Ltd. and in 2001, a R&D Engineer with 3COM Corporation. He is the Founder of Nova Greentech, Inc., an energy Internet company. He has authored more than 300 peer-reviewed articles in his areas of current research, which include large-scale dynamic complex system modeling and optimization, and its applications in the Internet and the energy Internet, especially in 5G sustainable powering. He is a Member of ACM. He was the Editor or the Guest Editor in many journals and was on TPCs of numerous international conferences.

YONG ZUO received the Ph.D. degree in communication engineering from the Chinese Academy of Sciences, Beijing, China, in 2012. He is currently an Associate Professor with the College of Electronic Science and Technology, National University of Defense Technology, Changsha, China.

His research interests include satellite communications, Internet of Things, integration of communication and navigation, edge computing, and artificial intelligence in wireless communication domain.