IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 April 2024, accepted 11 May 2024, date of publication 14 May 2024, date of current version 23 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3401016

RESEARCH ARTICLE

Deployment of Unmanned Aerial Vehicles in Next-Generation Wireless Communication Network Using Multi-Agent Reinforcement Learning

RAHUL SHARMA[1], SHAKTI RAJ CHOPRA[1], AKHIL GUPTA[1], (Senior Member, IEEE), RUPENDEEP KAUR[2], SUDEEP TANWAR[3], (Senior Member, IEEE), GIOVANNI PAU[4], (Senior Member, IEEE), GULSHAN SHARMA[5], FAYEZ ALQAHTANI[6], AND AMR TOLBA[7], (Senior Member, IEEE)

[1]School of Electronics and Electrical Engineering, Lovely Professional University, Phagwara, Punjab 144411, India
[2]Department of Electronics Technology, Guru Nanak Dev University, Amritsar, Punjab 143005, India
[3]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat 382481, India
[4]Faculty of Engineering and Architecture, Kore University of Enna, 94100 Enna, Italy
[5]Department of Electrical Engineering Technology, University of Johannesburg, Johannesburg 2006, South Africa
[6]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 12372, Saudi Arabia
[7]Department of Computer Science, Community College, King Saud University, Riyadh 11437, Saudi Arabia

Corresponding authors: Shakti Raj Chopra (shakti.chopra@lpu.co.in), Giovanni Pau (giovanni.pau@unikore.it), and Sudeep Tanwar (sudeep.tanwar@nirmauni.ac.in)

This work was supported in part by the Researchers Supporting Project through King Saud University, Riyadh, Saudi Arabia, under Grant RSPD2024R681; and in part by the Kore University of Enna, Enna, Italy.

ABSTRACT To address the challenges posed by a large number of disaster-waiver-affected users and the complexities of scaling centralized algorithms for rapidly restoring emergency communication services, the paper proposes a distributed intent-based optimization architecture based on multi-agent reinforcement learning. This approach aims to mitigate service discrepancies and dynamics among users. In the network feature layer, a distributed K-sums clustering algorithm considers variations in user services. Each UAV base station autonomously and minimally adjusts the local network structure based on user requirements. It selects user features from the cluster center as input states for the multi-agent reinforcement learning neural network. In the trajectory regulation layer, the paper introduces a multi-agent maximum entropy reinforcement learning (MASAC) algorithm. The UAV base station, acting as an intelligent node, governs its flight trajectory within the framework of ''distributed training – distributed execution.'' The paper incorporates techniques such as integrated learning and curriculum learning to enhance training stability and convergence speed. Simulation results demonstrate the effectiveness of our distributed K-sums clustering algorithm in terms of load efficiency and cluster balance, outperforming the traditional K-means algorithm. Additionally, the UAV base station trajectory control algorithm based on MASAC significantly reduces communication interruptions, enhances network spectral efficiency, and surpasses existing reinforcement learning methods.

INDEX TERMS Unmanned aerial vehicles, disaster, integrated learning, spectral efficiency.

I. INTRODUCTION

Following major natural disasters, ground-based communication infrastructure is often severely damaged, resulting in communication breakdowns and the loss of critical information, thereby jeopardizing the safety of affected individuals and complicating post-disaster rescue efforts. Unmanned Aerial Vehicles (UAVs), due to their swift deployment and adaptability, offer a viable solution by establishing Line of Sight (LoS) communication coverage in disaster-stricken

The associate editor coordinating the review of this manuscript and approving it for publication was Zihuai Lin.

areas. This approach holds significant potential for emergency communication [1]. As mobile Internet and Internet of Things (IoT) technologies have rapidly advanced, numerous digital devices and equipment have been deployed in emergency services, including but not limited to rescue operations and intelligent healthcare. Additionally, a plethora of sensors and auxiliary devices are now in place for continuous monitoring of disaster-stricken regions [1]. Consequently, the emergence of the 6G network has ushered in the demand for larger-scale, higher-density, and faster coverage in the realm of emergency communication [2]. Moreover, this network must address the challenges arising from the high dynamics and diverse service requirements resulting from large-scale user connectivity [3]. In response to the 6G landscape, the concept of an intelligent emergency communication network [4], [5] characterized by ''node intelligence and network simplification'' has emerged. By incorporating intelligent technologies, particularly the convergence of communication and computing [6], network nodes are equipped with intrinsic intelligence. This transformation leads to a simplified protocol structure within the network, promoting native simplicity. Furthermore, it facilitates on-demand, real-time adjustments in communication links and network configurations, driven by endogenous intelligence. This Intent-Driven Emergency Communication Network possesses the ability to dynamically adapt to user conditions, modify network deployments on the fly, and allocate network resources according to specific user service requirements.

Conventional non-intelligent emergency communication networks often rely on non-convex optimization techniques to enhance coverage performance. In such networks, coverage performance is heavily influenced by the real-time positions of UAV base stations relative to ground users. This necessitates solving the non-convex optimization problem related to the flight trajectory of UAV base stations. For instance, Chen et al. [7] developed a model for multi-user communication involving multiple UAV base stations and optimized the flight trajectories of these stations using iterative Gibb's sampling and block coordinate descent methods. This approach efficiently improved the network's maximum-minimum rate. Yin et al. [8] employed a continuous convex approximation method to jointly optimize the hover positions of ground clustering and multi-UAV base stations in large-scale ground user scenarios, thereby enhancing network spectral efficiency. The Zhang et al. [9] tackled power allocation and trajectory optimization issues for multiple UAV base stations, considering the communication characteristics and requirements of emergency communication scenarios. Their goal was to maximize the capacity of emergency communication networks. However, the aforementioned traditional non-intelligent coverage optimization methods rely on precise network environment state information (e.g., user locations, data sizes, channel conditions, etc.) as fixed parameters throughout the optimization process. As a result, these methods are primarily suitable for entirely static network scenarios where all network status information and service requirements of users in the future are known in advance. They are ill-suited to handling the dynamics and service variations of users in the aftermath of large-scale disasters.

Deep reinforcement learning is recognized as a pivotal technology for addressing network dynamics. UAV base stations equipped with deep reinforcement learning agents can adapt their flight trajectories based on real-time network conditions to maximize the network's long-term performance benefits. To derive the optimal coverage optimization strategy, the deep reinforcement learning agent undergoes iterative training and execution phases, which are crucial for adapting to the dynamic network environment and real-time regulation of UAV base station flight trajectories. Different approaches to the training and execution phases have led to various coverage optimization methods based on deep reinforcement learning. For instance, in [10], the deep reinforcement learning proximal strategy optimization (PPO) algorithm is employed, resulting in improved communication rates for single UAV base stations and reduced flight energy consumption. Zhang et al. [11] utilized the deep deterministic policy gradient (DDPG) algorithm to optimize the deployment of multiple UAV base stations without considering interference. However, when interference occurs between multiple UAV base stations, the single-agent learning environment becomes unstable, making it challenging for the algorithm to converge. To address these issues, Challita et al. [12] integrated game theory into the echo state network (ESN) and jointly optimized the flight trajectories of multiple UAV base stations. In contrast to the value function-based reinforcement learning method in [12], [13], and [14] employs the multi-agent deep deterministic strategy gradient (MADDPG) algorithm. This approach generalizes the action space using strategy gradients and can continuously output actions to accurately regulate UAV flight trajectories, avoiding the problem of dimension explosion [15]. However, as the scale of the emergency communication network grows, the input dimension of the MADDPG algorithm, based on the ''centralized training-distributed execution'' framework, increases significantly, leading to heightened learning complexity, reduced stability [16], and limited effectiveness in dealing with the coverage optimization challenges of large-scale post-disaster user scenarios and limited effectiveness in dealing with the coverage optimization challenges of large-scale post-disaster user scenarios.

After a disaster, site information should be quickly transmitted from the incident area to the rescue centre, with post-disaster communications able to perceive environmental and personnel information. Rebuilding the distributed intent-based coverage optimization architecture for wireless networks is, therefore, a practical way to serve a large number of post-disaster consumers. Effectively monitoring the fully mechanized mining face is completed when the source node gathers data about the mine catastrophe and utilizes a multi-hop routing data transmission mechanism to deliver the

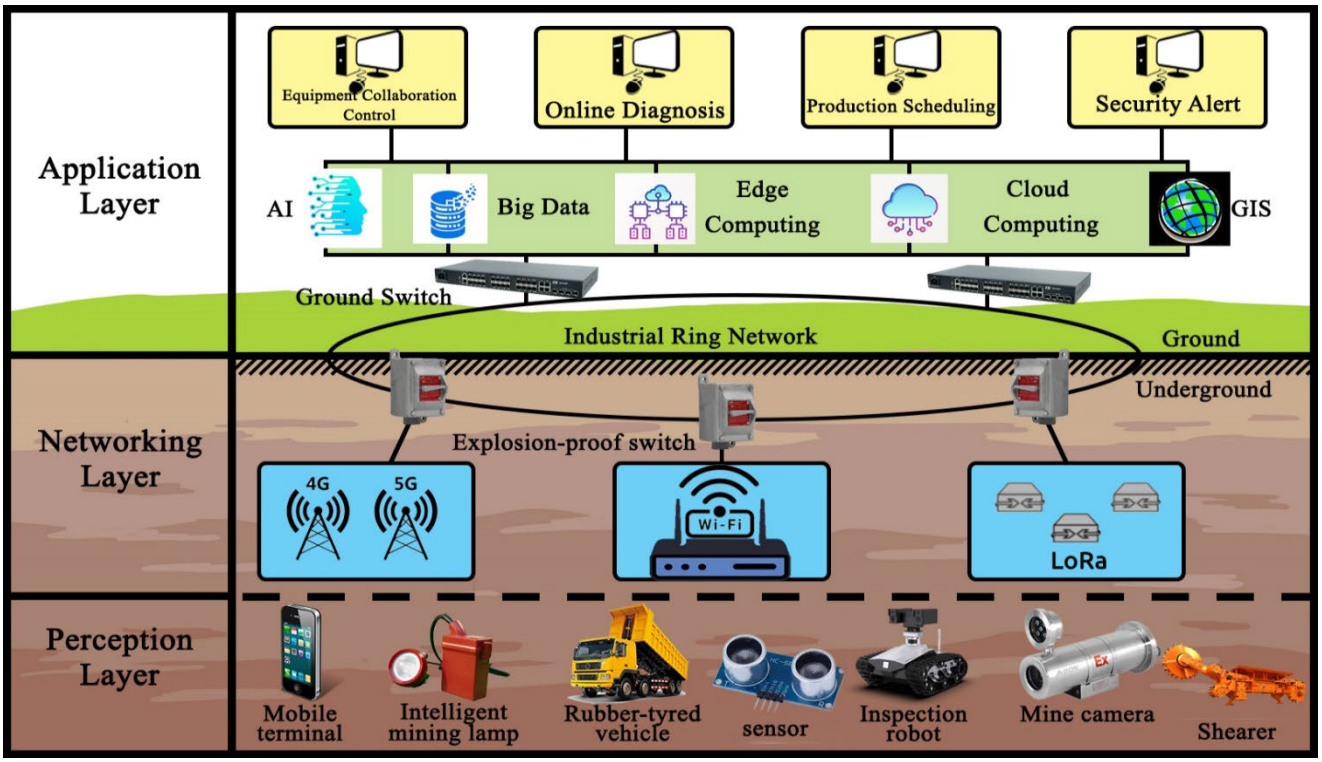

FIGURE 1. Distributed intent-based coverage optimization architecture for large-scale post-disaster users.

data packet to the sink node. In contrast to typical mining settings, nodes that survive in post-disaster dispersed networks possess copious amounts of energy that are not readily regenerated. Distributed intent-based coverage optimization design for several post-disaster users is shown in Figure 1. Post-disaster communications may typically resume by installing cables or additional communication equipment in accident tunnels. Post-disaster communications should also be able to sense personnel and environmental data and quickly relay site information from the incident site to the rescue centre.

To address the aforementioned challenges, this paper presents a distributed and intent-driven architecture for large-scale post-disaster user coverage optimization. In this architecture, the network feature layer adapts to the unique service requirements of large-scale post-disaster users by reconfiguring user clustering networks on demand. Multi-agent reinforcement learning technology is employed to empower each emergency UAV base station to intelligently and independently determine its flight trajectory, thereby enhancing the overall coverage performance of the emergency communications network. The key research contributions of this paper can be summarized as follows:

1). *Design of a Distributed Intent-Driven Large-Scale Post-Disaster User Coverage Optimization Architecture Based on Multi-Agent Reinforcement Learning:*

The architecture is specifically designed to leverage multi-agent reinforcement learning technology. In the feature extraction layer, it conducts distributed clustering of ground users using locally acquired network environment information. The resulting characteristic cluster center user information is then used as input states for the multi-agent reinforcement learning neural network. This enables the trajectory control layer to regulate the real-time flight trajectory of UAV base stations within a small-dimensional state space.

2). *Introduction of a Distributed K-SUMS Clustering Algorithm Tailored to User Service Differences:*

This paper proposes a distributed K-SUMS clustering algorithm to capture the post-disaster user dynamics in large-scale disasters while considering their diverse service needs. The algorithm employs Bayesian inference to facilitate online learning of user service differences, yielding priority coefficients for user transmissions. Subsequently, UAV base stations perform distributed clustering based on these priority coefficients and local user load information, leading to the identification of cluster center users. Compared to traditional clustering methods, the distributed K-SUMS clustering algorithm demonstrates improved performance in terms of load efficiency and inter-cluster balance.

3). *Introduction of the Multi-Agent Soft Actor-Critic (MASAC) Algorithm for Distributed UAV Base Station Trajectory Control:*

This paper puts forth the MASAC algorithm, a multi-agent reinforcement learning technique that empowers individual UAV base stations to autonomously adjust their flight trajectories. Employing the ''distributed training and distributed execution'' framework, MASAC integrates

maximum entropy theory, ensemble learning, and curriculum learning techniques. This integration addresses the issues faced by existing multi-agent deep reinforcement learning methods, which often suffer from instability and are significantly impacted by disaster scenarios. The MASAC algorithm effectively reduces the frequency of communication interruptions in emergency communication networks and enhances network spectrum utilization efficiency.

II. SYSTEM MODEL AND ARCHITECTURE DESIGN

Figure 2 demonstrates two sections, Space and Ground section in which the Space section facilitates communication between Earth stations or between Earth stations and spacecraft, communication satellites serve as the primary means of receiving and transmitting signals from satellite communication Earth stations and Ground section enable user connection, the ground segment consists of a tracking, telemetry, and command station (TT&C), satellite transponder, gateway station, and satellite control center (SCC). The user sector is primarily made up of different terminal user devices, including as handheld terminals, mobile terminals installed on vehicles, ships, and aircraft, and Very Small Antenna Terminal (VSAT) stations. It also includes a variety of satellite communication-based applications and services.

Satellite communication is highly suitable for emergency communications because it is not constrained by ground conditions and offers a wide range of communication coverage, large capacity, high reliability, long transmission distance, independent communication ability, and strong resistance to damage.

Figure. 2 observes a large-scale dynamic population of diverse ground users within the disaster-affected area. To cater to their communication needs, multiple UAV base stations are strategically deployed to form an emergency communication network. Let's assume there are N users in the disaster-affected region, and M UAV base stations have been deployed. The users are logically grouped into M distinct clusters, each serviced by a specific UAV base station to reestablish communication services. In each user cluster, there exists a designated cluster center directly connected to the UAV base station serving that cluster. Communication from other users within the cluster is routed through this central user. The UAV base station is designated as M, while the user is designated as N.

Within the expansive emergency communication network, the utilization of cluster center users for aggregating and forwarding information offers distinct advantages in three key areas: processing capacity, energy efficiency, and interference mitigation. Firstly, considering the limited processing capacity of UAV base stations, user clustering reduces the number of users directly connected to these stations. This, in turn, effectively reduces the dimensionality of the neural network, preventing network paralysis. Secondly, user clustering leads to a decrease in the number of users directly connected to UAV base stations. This reduction results in lower communication and computing energy consumption for the UAV base stations, thus extending their continuous operational duration. Lastly, the reduction in the number of air-ground communication links, facilitated by user clustering, minimizes interference between air-ground communication clusters. This, in turn, enhances the overall communication capability of the network.

A. USER MODEL

In the actual and intricate emergency communication network environment, it's evident that large-scale post-disaster users exhibit substantial dynamism and variations in service requirements. This dynamism is primarily observed in the real-time fluctuations in the users' positions and the temporal randomness of their activation states. When a user becomes active at a specific moment, a new data transfer task emerges. The activation state of user I follows a Beta distribution within the time interval $t\epsilon[0,T]$, where

$$f_i(t) = \frac{t^{k_1-1}(T-t)^{k_2-1}}{T^{k_1+k_2-1}B(k_1+k_2)} \tag{1}$$

$$B(k_1+k_2) = \int_0^1 t^{k_1-1}(1-t)^{k_2-1}dt \tag{2}$$

The parameters k_1 and k_2 define the characteristics of the Beta distribution. Importantly, a user's activation status is contingent upon the presence of a new transfer task. Even when a user is inactive, they can still complete the transfer of any remaining data from previous tasks and may subsequently be designated as the cluster center user. Upon being assigned as the cluster center user, they bear the responsibility of relaying information for all users in the cluster and typically require higher transmission power. Since this article focuses on coverage optimization to restore communication for a large number of users, it does not delve into energy balancing for users. The central consideration is the disparity in information due to diverse service types and task requirements, particularly variations in the data sizes that users need to transmit. For an activated user i at time t, the data size of their new transfer task, denoted as di(t), follows a Gaussian distribution [17].

$$f_d(d_i(t)) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(\frac{(d_i(t)-\mu_i)^2}{2\sigma_i^2}\right) \tag{3}$$

Here, μ_i and σ_i represent constant values that define the mean and standard deviation, respectively, characterizing the transfer task size for user I and their specific service type. The value of $d_i(t)$ can vary over time due to semantic changes in the transmission task.

B. GROUND-BASED TRANSMISSION MODEL

In this large-scale disaster scenario, ground users are grouped into M clusters, aligning with the number of UAV base stations. Each user initiates data transmission to the designated cluster center user, and subsequently, the data is routed to the UAV base station through the forwarding of the cluster center user. The communication between user i and cluster

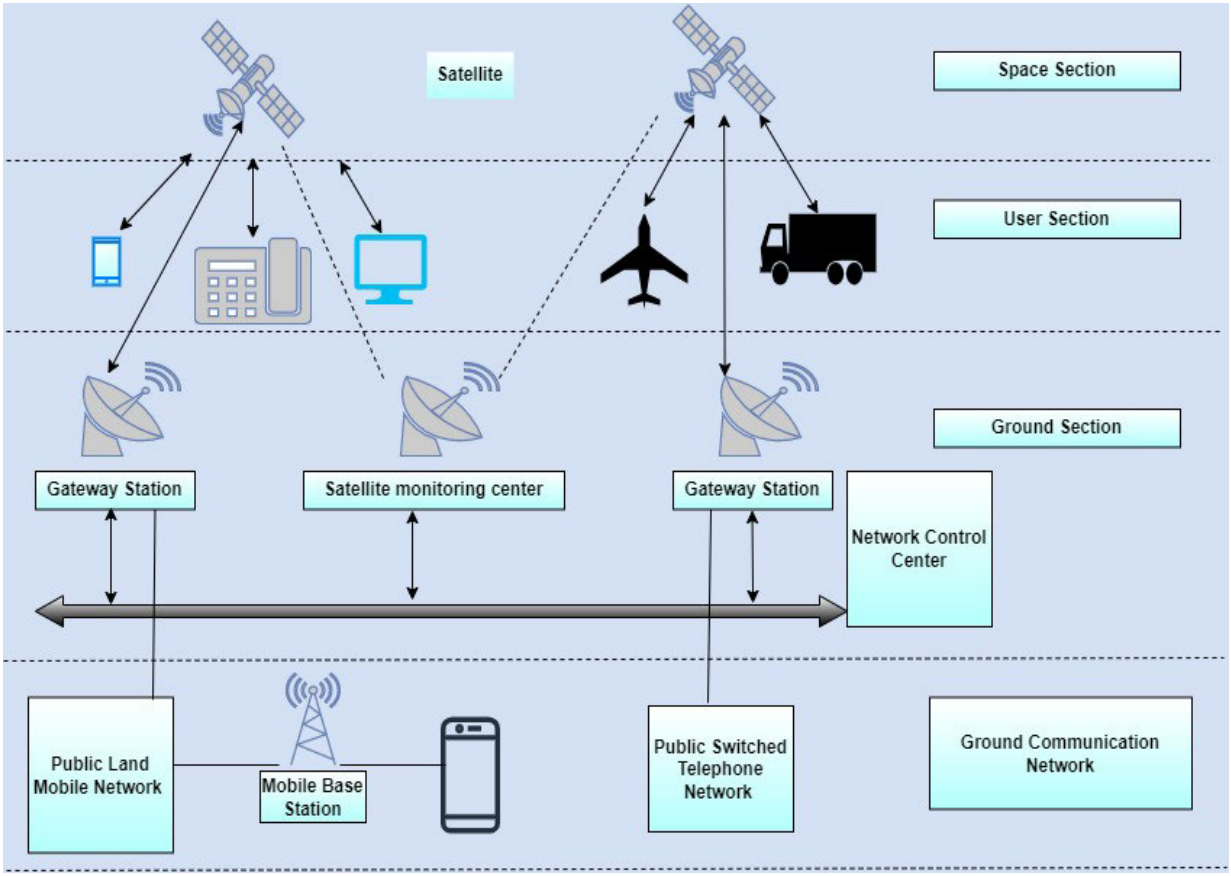

FIGURE 2. Emergency communication network system model.

center user μ_i operates on ground-to-ground communication links within the sub-6 GHz band, where non-line-of-sight (NloS) conditions significantly impact the wireless link. The path loss, following the Rayleigh fading channel model [18], is expressed as:

$$L_{i,u_i}^{ground}\,(dB) = 37.6log\left\|p_i - p_{u_i}\right\| + 21logf_c^{ground} + 58.8 \tag{4}$$

In this equation, f_c^{ground} represents the central frequency used for terrestrial communications, p_i and p_{u_i} are the positions of user I and cluster center user u_i, $\|p_i - p_{u_i}\|$ denotes the Euclidean distance between these two locations. The coefficients 37.6 and 21 account for the distance attenuation and frequency attenuation factors in the path loss model for non-high-rise urban or suburban scenarios. The constant term 58.8 is an additional path loss constant that considers the height difference between users. Notably, due to the considerable distance between cluster users, interference between clusters can be effectively minimized through appropriate spectrum resource allocation techniques. However, this paper does not delve into spectrum resource allocation. The Signal-to-Interference plus Noise Ratio (SINR) for the communication link between user i and cluster center user u_i within the cluster can be expressed as

$$SINR_{i,u_i}^{ground} = \frac{P_1 G_{i,u_i}^{ground}}{N_0} \tag{5}$$

In this equation, P_1 represents the transmit power of user i, and ground, G_{i,u_i}^{ground} represents the channel gain between user i and the cluster center user u_i. N_0 represents the noise power.

The channel gain ground, G_{i,u_i}^{ground} is influenced by path loss and can be described as follows:

$$P_1 G_{i,u_i}^{ground}\,(dB) = P_1\,(dB) - L_{i,u_i}^{ground}(dB) \tag{6}$$

The spectral efficiency of data transmission by user i at time t can be represented as follows:

$$R_{i,u_i}\,(t) = lb\left(1 + SINR_{i,u_i}^{ground}\,(t)\right) \tag{7}$$

The overall transfer task size of the user i at time t is denoted as $D_i(t)$ and encompasses the remaining task from time $(t-1)$, denoted as $D_i\,(t-1)$, and the new task size $D_i(t)$ at time t. If there's no remaining task at the start, i.e., $D_i\,(0)0 = 0$, then

$$D_i\,(t) = max\left(0, D_i\,(t-1) - n_i\,(t-1)\,BR_{i,u_i}\,(t-1) + d_i\,(t)\right) \tag{8}$$

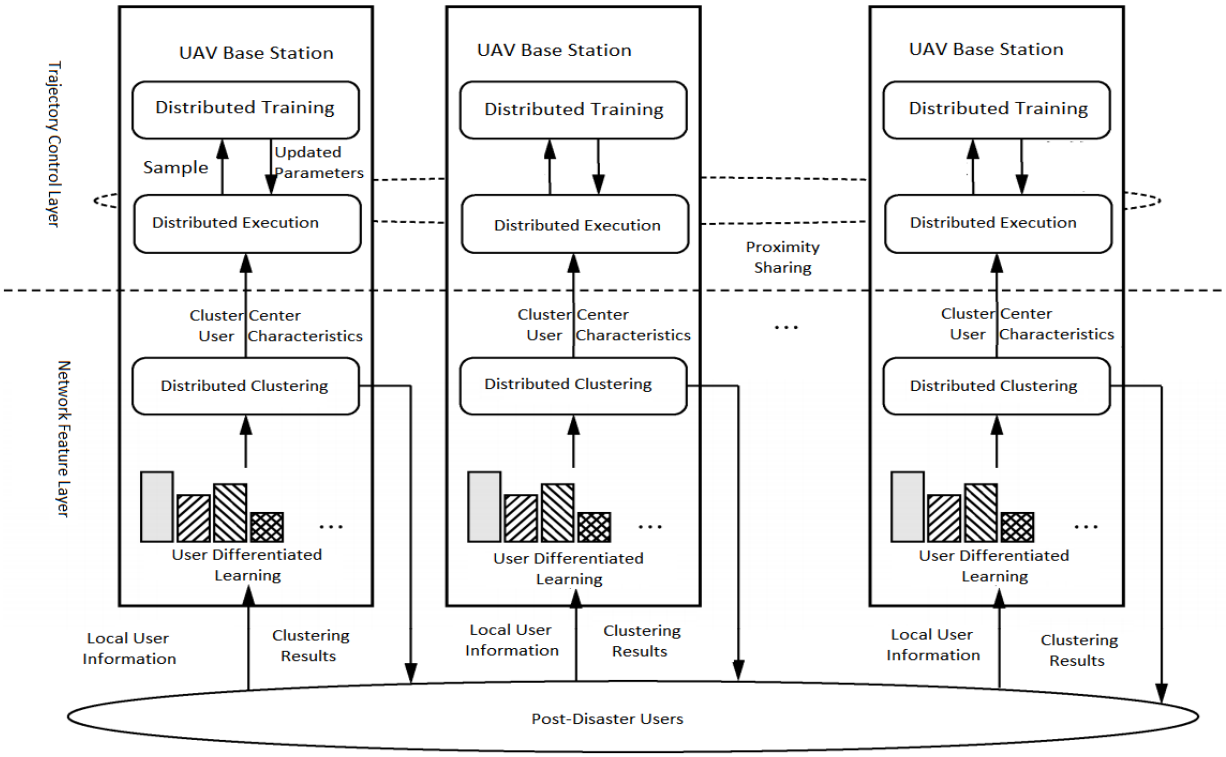

FIGURE 3. Architecture for distributed intent-based coverage optimization in large-scale post-disaster user scenario.

Here, B signifies the bandwidth of the ground resource block, and $i(t)$ n represents the number of resource blocks allocated to the user, which is determined by the total transmission task size and spectral efficiency at time t.

Where N_C represents the threshold for resource block load, which prevents users from excessively consuming spectrum resources due to low spectral efficiency. The average load efficiency evaluation index is defined as

$$n_{i,u_i}(t) = min\left(N_c, \left[\frac{D_i(t)}{R_{i,u_i}(t)}\right]\right) \tag{9}$$

$$\eta = \frac{1}{TN}\sum_{t=0}^{T}\sum_{i=0}^{N}\frac{BR_{i,u_i}(t)}{n_{i,u_i}(t)} \tag{10}$$

The efficiency of load averaging, denoted as η, serves as an effective metric for evaluating the quality of ground clustering results across various scenarios characterized by different user dynamics and information disparities.

C. FOR MODEL TRANSMISSION BETWEEN AIR AND GROUND

Communication between the emergency UAV base station and the cluster center user occurs through air-to-ground communication links in the sub-6 GHz band, where Line of Sight (LoS) conditions prevail, significantly influencing the wireless link. The average path loss between UAV base station J and cluster center user u_j is formulated as follows:

$$L_{j,u_j}^{air}(dB) = 20log\left(\frac{4\pi f_c^{air}\left\|p_j - p_{u_j}\right\|}{c}\right) + \eta_{LoS} \tag{11}$$

Here, f_c^{air} denotes the central frequency of air-ground communication, p_j signifies the position of the drone base station, c represents the speed of light, and η_{LoS} denotes the additional spatial propagation loss for Line of Sight (LoS) and is treated as a constant. It's important to note that cluster center users may introduce interference to other UAV base stations, and the Signal-to-Interference-Noise Ratio (SINR) for the communication link between UAV base station J and cluster center user u_j for the service is given as follows:

$$SINR_j^{air} = \frac{P_2 G_{j,u_j}^{air}}{N_0 + \sum_{j'\neq j, j'\epsilon M} P_2 G_{j,u_{j'}}^{air}} \tag{12}$$

Here, P_2 represents the transmit power of the cluster center user and $L_{j,uj}^{air}$ represents the channel gain between the UAV base station j and the cluster center user u_j. The channel gain $G_{j,uj}^{air}$ is influenced by path loss and can be expressed as,

$$P_2 G_{j,u_j}^{air}(dB) = P_2(dB) - L_{j,u_j}^{air}(dB) \tag{13}$$

The Doppler effect induced by drone movement can be effectively compensated for using existing technologies, such as phase-locked loop technology. The spectral efficiency of

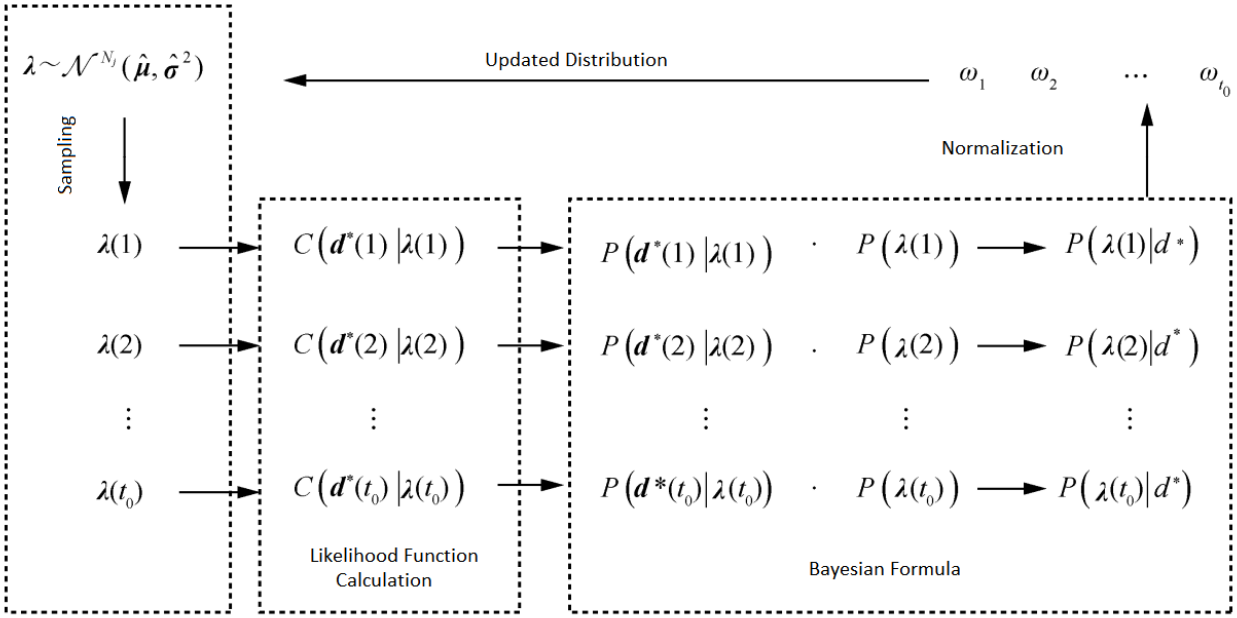

FIGURE 4. Bayesian inference flow.

UAV base station j can be represented as

$$R_j(t) = lb\left(1 + SINR_j^{air}(t)\right) \tag{14}$$

The average spectral efficiency of an emergency communication network can be expressed as:

$$R(t) = \sum_{j \epsilon M} lb\left(1 + SINR_j^{air}(t)\right) \tag{15}$$

In this study, the optimization goal is set as the average spectral efficiency defined in Equation (15). The optimization problem is formulated while considering the constraints imposed by the maximum flight speed limit, flight safety limitations, and communication interruption limits of UAV base stations.

$$\begin{aligned} \text{OP:} \max_{\mathrm{p_j(t)}, \mathrm{j}\in\mathrm{M}, t=1,\ldots,\mathrm{T}} & \sum_{\mathrm{t=1}}^{\mathrm{T}} \sum_{\mathrm{j}\in\mathrm{M}} \mathrm{lb}\left(1 + \mathrm{SINR_j^{air}(t)}\right) \\ \text{s.t} \quad & C_1 : \left\|p_j(t) - p_j(t+1)\right\| \le V_{max}\Delta t, \forall j \in M \\ & C_2 : \left\|p_j(t) - p_j(t)\right\| < 0, \quad \forall j \ne j' \in M \\ & C_3 : P_{outage}(t) \le P_{outage}^{\max} \end{aligned} \tag{16}$$

In the optimization problem, the communication interruption probability (P_{outage}) and the maximum communication interruption probability limit (max P_{outage}) at time-t are represented by the outage function $P(t)$.

The average spectral efficiency of the emergency communication network in this optimization problem is determined by the signal-to-noise ratio between each UAV base station and the cluster center users. Since air-ground communication mainly involves a direct path, the signal-to-noise ratio is primarily influenced by the distance between the two. Additionally, constraint C_3, which relates to communication interruption, is closely linked to the selection of ground user clustering and cluster center users. Therefore, the trajectory adjustment in a large-scale multi-UAV emergency communication network depends on the outcomes of ground user clustering, and the flight trajectory adapts to the dynamic changes in user selections for the cluster center.

D. OPTIMIZATION SCHEME OVERRIDES

Based on the previously discussed user and communication models, the average spectral efficiency (RT) of the emergency communication network depends on various factors such as the positions of UAV base stations (p_j), the locations of cluster center users (p_u), and the results of ground user clustering. To address this, Figure 3 depicts the architecture, which consists of two layers: the network feature layer and the trajectory control layer. This distributed intent-based large-scale post-disaster user coverage optimization technology. Compared to the traditional end-to-end coverage optimization structure, the hierarchical coverage optimization structure proposed in this paper offers several advantages:

Reduction in input dimension: By reducing the input dimension of the reinforcement learning state at the UAV base station, it is possible to simplify problem training by reducing the deep neural network's scale.

Hierarchical design: Through this hierarchical approach, separate air communication optimization and ground communication optimization, make it easier to adjust performance and parameters in practical engineering applications. This hierarchical design aligns with the typical

implementation of deep reinforcement learning algorithms in various industries. Specifically, each UAV base station is equipped with a distributed computing terminal that serves the layered optimization architecture mentioned above. In the network feature layer, the UAV base station leverages locally acquired network status information to accommodate the service differences among large-scale post-disaster users. It autonomously groups local users based on this information and selects user features from the cluster center as the input state for multi-agent reinforcement learning. In the trajectory control layer, multi-agent reinforcement learning technology is employed to address the state input for time series dynamics, with UAV base stations autonomously optimizing their flight trajectories within the framework of distributed training and execution. This optimization aims to reduce communication interruptions and maximize network spectral efficiency. It's worth noting that, in each timeframe, along with user information, the characteristics of the cluster center user, who relays information in the transmission process, are aggregated. These characteristics are also required as input for reinforcement learning and are transmitted to the UAV base station as auxiliary communication overhead.

III. NETWORK FEATURE LAYER-GROUND USER CLUSTERING

In the network feature layer, the process of ground user clustering and the selection of cluster center users are essential to address the varying service requirements among the large-scale user population. This section introduces a user differentiation learning algorithm based on Bayesian inference to address this challenge. As obtaining information about all the large-scale users can be challenging for UAV base stations, the paper also presents a distributed K-SUMS clustering algorithm that considers these user differences. This algorithm results in clustering outcomes characterized by improved load efficiency and a more balanced distribution of users across clusters.

A. USER DIFFERENTIATED LEARNING

"Bayesian Inference" is a statistical machine learning method that establishes a connection between an observer and an estimator using Bayesian formulas [19]. In the process of user differentiation learning, the UAV base station can acquire the new task size of the user's latest activation at time T_0 as an observation d_i^* and estimate the user's priority parameter λ_i. In this paper, the priority parameter λ_i represents a numerical representation of the average traffic demand of user i, considering information differences, to allocate higher-quality spectrum resources to users with higher priority. λ_i follows a Gaussian distribution with mean μ_i and variance σ^2.

Suppose the number of local user's observable by UAV base station j is Nj, represented by the set Nj, and the definition vector is x = (x, y, z), where d^* is the observation vector, λ is the estimation vector, and μ and σ^2 are parameter vectors.

The Bayesian inference process is illustrated in Figure 4. Initially, the estimation vector λ, which has the same number of dimensions as the observation vector, is obtained through sampling from the prior distribution $\lambda \sim N(\mu, \sigma^2)$, where $P(\lambda)$ is the prior probability distribution. Then, using the vectors d^* and λ, the loss function is computed as follows:

$$C\left(d^* \mid \lambda\right) = -\frac{1}{N_j} \sum_{i=0}^{N_j} \frac{\left(d_i^* - \lambda_i\right)^2}{d_i^* \lambda_i} \tag{17}$$

The likelihood function can be derived by normalizing the loss function with respect to the estimation vector λ given the observation vector d^*, as follows:

$-C(d|\ \lambda)$ where $Z(d)$ is the normalization constant.

$$P\left(d^* \mid \lambda\right) = \frac{e^{-C\left(d^* \mid \lambda\right)}}{\int e^{-C\left(d^* \mid \lambda\right)} dd^*} \tag{18}$$

$$P\left(\lambda \mid d^*\right) \propto P\left(d^* \mid \lambda\right) P(\lambda) \tag{19}$$

Based on equation (19), the product of the prior probability and likelihood function is normalized to obtain the posterior probability ω for the estimated vector λ. Consequently, the mean and variance of the prior distribution are updated

$$\hat{\mu} = \sum_{t=1}^{t_0} \lambda\ (t)\omega \tag{20}$$

$$\hat{\sigma}^2 = \sum_{t=1}^{t_0} \left(\lambda\ (t) - \hat{\mu}\right)^2 \omega \tag{21}$$

Algorithm 1 *User Differentiation Learning Algorithm Based on Bayesian Inference*

Input:
Observation vector,d∗
Parameter vectors to be optimized,μ and 2σ
Output:
Optimized parameter vectors,μ and 2σ
1. Initialize the priority parameters for t_0 group users from the prior distribution:λ (1),λ(2), . . . ,$\lambda(t_0) \sim N(\mu, 2\sigma)$.
2. For $t = 1$ tot_0 :
Calculate the loss function using Equation (17):
3.$C(t, d*, \lambda)$ to characterize the gap between the observation vector and the sampled priority parameter vector.
4. Normalize the loss function using Equation (18) to obtain the likelihood function:$P(t|d*, \lambda)$.
5. Use Equation (19) to calculate the product of the loss function and the likelihood function based on Bayesian inference.
6. End for
7. Multiply all loss functions with likelihood functions and normalize to obtain the posterior probability : $\omega(t), t = 0, 1, 2, \ldots, t\_0$.
8. Update the posterior distribution using Equation (20) and Equation (21) to obtain parameters μ and σ^2.

Algorithm 1 computes the priority parameter λ for each user, allowing for differentiated communication services during clustering. Prioritizing higher λ users for spectral efficiency can effectively reduce the network spectrum resource load.

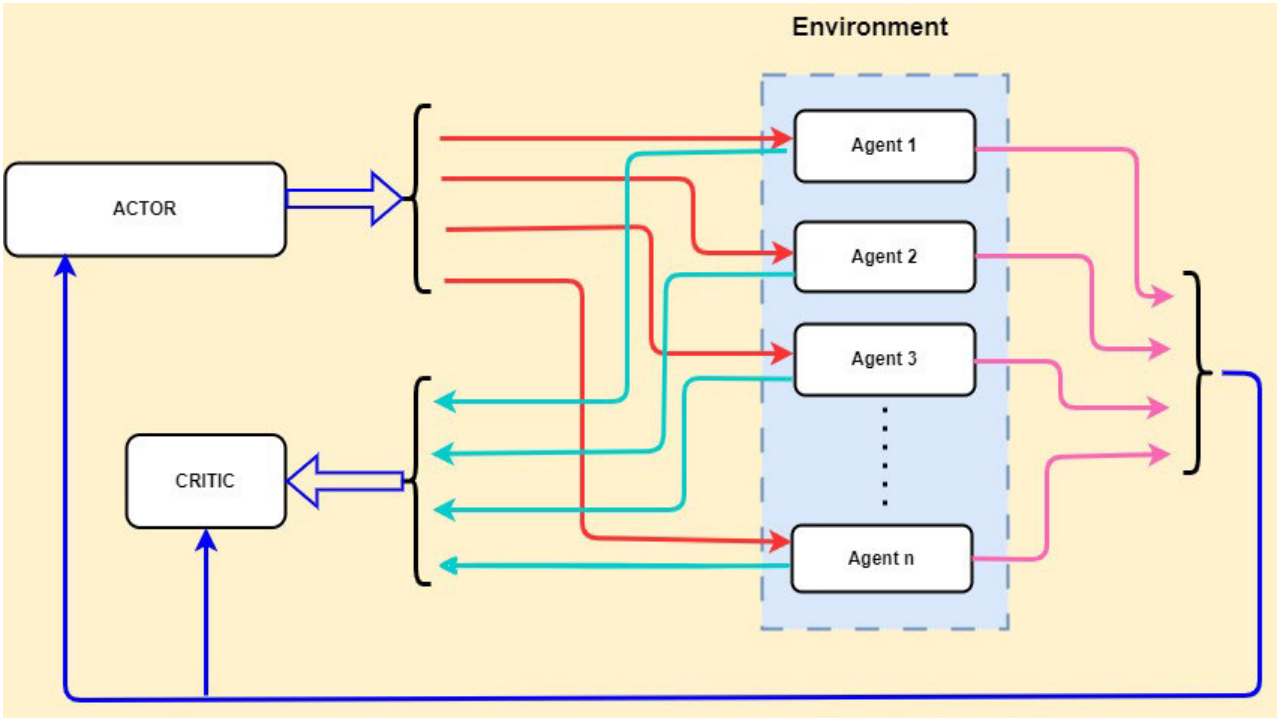

FIGURE 5. Basic multi-agent reinforcement learning MASAC agent structure.

B. USERS CLUSTERING

The k-sums algorithm is chosen over traditional clustering methods like k-means and spectral clustering due to its lower computational complexity, which is O(NM), making it efficient even when users and cluster centers change rapidly. Moreover, the k-sums algorithm effectively reduces intra-cluster distances, improving the balance of users between clusters. These characteristics are crucial for optimizing average spectral efficiency and load balancing in an emergency communication network. In general, the k-sums algorithm is well-suited to handle the dynamic and diverse nature of users in the aftermath of large-scale disasters. The general matrix expression for clustering algorithms can be expressed as The matrix notation used in the clustering algorithm is as follows:

$$\min_{y\in R^{N\times M}} \mathrm{Tr}\left(\left(Y^{T}Y\right)^{-\frac{1}{2}} Y^{T}GY\left(Y^{T}Y\right)^{-\frac{1}{2}}\right) \tag{22}$$

The Tr () operator represents the trace operation of the matrix.

The matrix notation used in the clustering algorithm is as follows:

Matrix Y represents the cluster assignment matrix with dimensions N×M. When user i belongs to service cluster J of the UAV base station, the element is set to 1 ($y_{i,j} = 1$); otherwise, it's set to 0 ($y_{i,j} = 0$).

Matrix G is the cluster kernel matrix, and its definition varies depending on the clustering algorithm. The k-sums algorithm involves neighbor dissimilarity measures between nodes. Elements g_{ii} represent the dissimilarity between user i_1 and user i_2. The smaller the dissimilarity, the larger the value of g_i, and only one element in each row (N_j) is set to 1, representing the smallest dissimilarity, while the other elements are replaced by a maximum dissimilarity constant.

To ensure the clustering results' balance, the k-sums algorithm introduces the constraint $Y^{T}Y=nI$ to equation (22), where I is the identity matrix, and n is an arbitrary constant. Equation (22) can be transformed into:

$$\begin{aligned} &\min_{y\in R^{N\times M}} \mathrm{Tr}(Y^{T}GY) \\ &\text{s.t.} Y^{T}Y = \bar{n}I \end{aligned} \tag{23}$$

In the case of large-scale post-disaster users as shown in Figure 8, obtaining information about all users for a single UAV base station is challenging. Therefore, calculating dissimilarity measures between all global users is infeasible. Adopting a centralized clustering approach in this scenario would result in significant communication overhead for user information. To address this issue, this paper introduces a distributed K-Sums clustering algorithm, allowing UAV base stations to perform clustering using only locally observed information from large-scale post-disaster users.

The clustered kernel matrix G in the distributed K-Sums algorithm is defined based on the proximity dissimilarity measure of observable users. The dimension of the clustered kernel matrix for UAV base station j is represented as $N_j \times N$, where N_j is the number of users observable by UAV base station j. The dissimilarity measure between users is calculated as the product of the number of load resource blocks required to transmit from user i_1 to user i_2 at the current moment

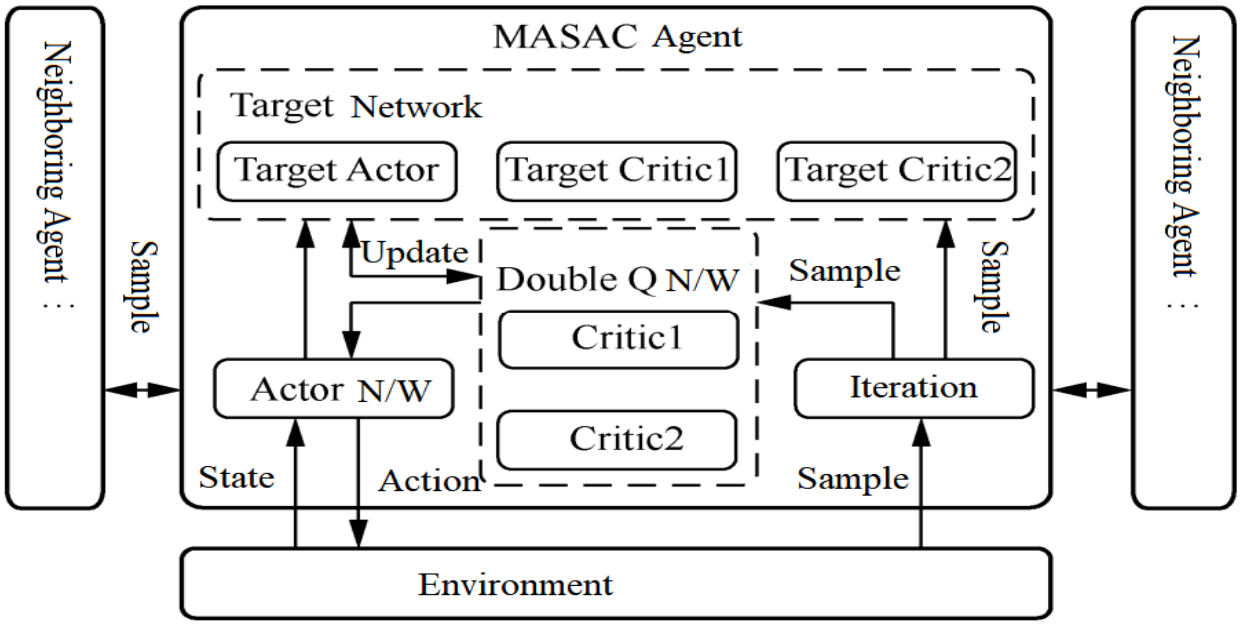

FIGURE 6. Detailed block diagram of multi-agent reinforcement learning MASAC agent structure.

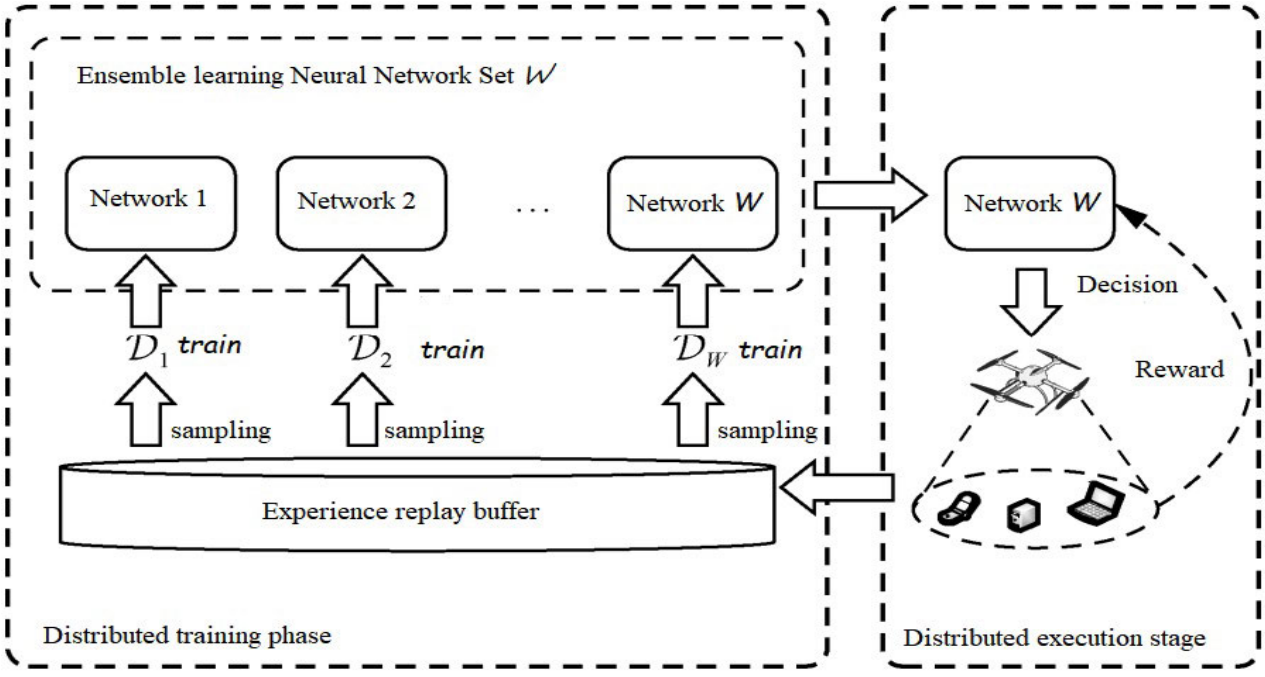

FIGURE 7. Implementation architecture of stable convergence technology based on ensemble learning.

$(n_{i1,i2})$ and the user's priority parameter λ_i. In mathematical terms, this dissimilarity measure is represented as:

$$g_{i_1,i_2} = n_{i_1,i_2}\lambda_{i_1} \tag{24}$$

This design aims to consider both the instantaneous and long-term characteristics of user transmission information traffic requirements. It allocates load resource blocks to users based on their information differences, providing better resource blocks to users with higher business needs. This approach effectively reduces the probability that high-priority user communication cannot be covered under limited load. It's important to note that the design of the cluster kernel matrix in this paper primarily focuses on differences in user traffic demand. If you need to consider business differences

arising from variations in other communication requirements, you would need to redefine the physical meaning of the elements in the cluster kernel matrix to accommodate those differences.

For each UAV base station, the distributed K-Sums clustering algorithm only needs to obtain the user cluster it serves. Therefore, it defines the local portion of the cluster identification matrix as $Yp \subseteq N_j \times 2$, where $y_{i,0}$ indicates whether user i is in the user cluster N_j served by the UAV base station. To ensure the balance of user clustering results and meet the conditions of equation (23), the elements of the matrix Y_p must satisfy:

$$y_{i,0} = \begin{cases} 1, & i \in N_j \\ 0, & i \notin N_j \end{cases} \tag{25}$$

$$y_{i,1} = \begin{cases} 1, & i \in N_j \\ \sqrt{\dfrac{N}{M(M-1)}}, & i \notin N_j \end{cases} \tag{26}$$

The local partial cluster identification matrix must satisfy the constraint $Y^T Y = nI$ of the global cluster identification matrix. Additionally, the number of observable users N_j for the UAV base station needs to be greater than the average number of users serviced by UAV base stations, i.e., $N_j > M$. Similar to the row iteration method used in the K-Sums algorithm [20], the local partial cluster of each user is optimized in sequence to identify the row vector, which is represented as $0 \leq y_i \leq 1$ for each row vector. The problem equation (23) can then be transformed into:

$$\min_{y_i} \mathrm{Tr}\left(Y_p^T \mathrm{GY}\right) \Longleftrightarrow \min_{y_i} y_i^T \tilde{Y}_p^T g_i \tag{27}$$

The distributed K-Sums clustering algorithm, which considers user differences, is presented in Algorithm 2:

Algorithm 2 Distributed K-Sums Clustering Algorithm with User Differences

Input: User dissimilarity metric matrix G, local part cluster identification matrix pY Output: Local part cluster identification matrix Yp after optimization

1. Initialize Yp and pY such that they are different from each other.

2. While $p \neq pY$*:*

i. $p \leftarrow pY$

ii. For each j in [1, N]:

a) *Perform row-wise optimization of the cluster identification matrix Yp according to Equation (27) to obtain the optimized yi.*

iii. End for

3. End while

By calculating the result Y_p from Algorithm 2, filter the users for whom $y_{i,0} = 1$ as the users serviced by UAV base station j, and select the user with the least similarity measure as the cluster center user, i.e., the user with the smallest $g_{i,i}$ value.

$$\min_{i_2 \in N_j, y_{i_2,0}=1} \sum_{i_1 \in N_j} g_{i_1,i_2} \tag{28}$$

Based on the distinctive information of cluster center users, UAV base stations can dynamically adjust their flight trajectories in real time to optimize coverage for ground users. This aspect will be explored in greater detail in Section III of this paper.

C. COMPLEXITY ANALYSIS

The standard k-means algorithm requires iterative assignments of users to the nearest cluster center, recalculating cluster centers for each user cluster, and thus calculating distances between each user and all cluster center users with a complexity of O(NM). However, the standard k-means algorithm has a limited scope of applicability, as it can only handle linearly separable data and is highly sensitive to initialization. The enhanced k-means algorithm first nonlinearly maps the input data to a higher-dimensional space to accommodate non-linearly separable data types and then performs the k-means algorithm with a computational complexity of O(2N).

The spectral clustering algorithm, on the other hand, leverages the nearest neighbor map of users for analysis, making it capable of processing non-linearly separable data with superior clustering performance. However, due to the initial construction of the proximity map and subsequent spectral decomposition operations, its computational complexity is high, reaching $O(N^2M)$. In contrast, the clustered kernel matrix of the k-SUMS algorithm utilizes neighbor dissimilarity measures, and most of the values in g_i are constant. Employing the row iterative optimization method as described in equation (27), the complexity is approximately O(M), resulting in an overall computational complexity of O(NM).

Moreover, to learn the business differences of users online, Bayesian inference algorithms require t0 steps to compute both the loss function $(C(t)|(d*(t), \lambda))$ and the likelihood function $(P(t)|(d*(t), \lambda))$, where the computational complexity of the loss function is associated with the number of locally observable users, N_j. Consequently, the computational complexity of the user differentiation learning algorithm based on Bayesian inference is $O(t_0 N_j)$. In summary, the network feature layer, which encompasses the entire ground user clustering, accounting for user differences, has a complexity of $O(t_0 N_j)$.

IV. CONTROL OF UAV TRAJECTORY AT THE BASE STATION LEVEL

The conventional approach to optimizing UAV base station trajectories is inadequate for addressing the dynamic and long-term aspects of large-scale user scenarios. Simultaneously, single-agent reinforcement learning methods struggle to adapt to the unstable learning conditions arising from multiple UAV base stations. To tackle these challenges

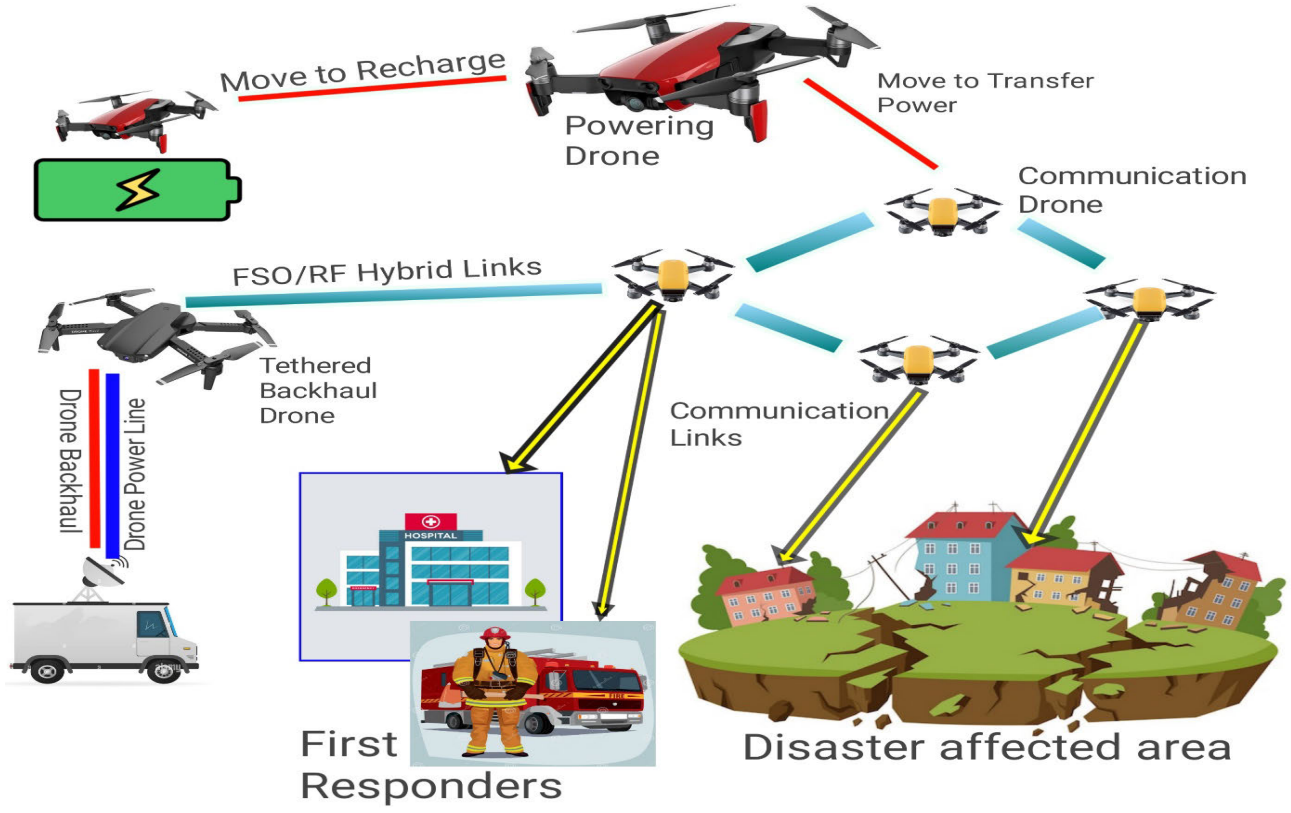

FIGURE 8. Overall process optimized for distributed coverage for large-scale post-disaster users.

effectively, introduce a multi-agent reinforcement learning-based optimization approach. This method makes intelligent decisions regarding flight trajectories by considering the current state of the network environment. This research, present the Multi-Agent Soft Actor-Critic (MASAC) algorithm, which offers superior convergence and stability compared to the existing Multi-Agent Deep Deterministic Policy Gradients (MADDPG) algorithm.

A. DESIGN OF A MULTI-AGENT REINFORCEMENT LEARNING-BASED DISTRIBUTED REGULATION AND CONTROL SYSTEM FOR UAV BASE STATIONS

In addressing the coverage optimization challenge for a large-scale post-disaster user scenario, Section I introduces a distributed intent-based coverage optimization framework. Within this architecture, the network feature layer assumes responsibility for clustering the extensive ground user population, selecting key user feature data from the cluster centers, and serving as the input layer for the trajectory control module in the multi-agent reinforcement learning system.

The trajectory regulation layer employs a multi-agent deep reinforcement learning approach, utilizing the Markov decision process to remodel the trajectory regulation problem. This transformation turns the global optimization problem into a series of reinforcement learning optimization objectives, focusing on maximizing network spectral efficiency over time. The design leverages reward functions and value functions to iteratively adjust the flight trajectory of the UAV base station.

Consequently, the distributed regulation design for UAV base stations based on multi-agent reinforcement learning operates as follows:

State: Each UAV base station extracts specific observable information as an input state. This information includes:

- The coordinates of the UAV base station itself.
- Two-dimensional relative positioning with respect to the ground cluster's central user.
- Signal-to-noise ratio of user information received by the cluster center.
- Three-dimensional relative positioning with neighboring drones (Mj).

Action: Considering the UAV base station's freedom of movement in three-dimensional space, its output actions are characterized by its speed in three directions: x-axis, y-axis, and z-axis.

Reward: The reward function comprises three components:

- Flight safety penalty value.
- Communication interruption penalty value.
- Spectrum efficiency reward value.

These components work together to guide the optimization process.

$$r_j = R_j(t) - \xi_{collision} I^j_{collision} - N_j P^j_{outage}(t)\, \xi_{outage} I^j_{outage} \tag{29}$$

In the given context, where $P^j_{outage}(t)$ represents the instantaneous communication interruption probability and $R_j(t)$ signifies the instantaneous network spectral efficiency, $\xi_{\text{collision}}$ and ξ_{outage} serve as constants for security and communication penalties when the UAV base station "j" encounters collisions or exits the specified area. When I_{outage} equals 1, it diminishes the reward function's magnitude in response to these events during the multi-agent reinforcement learning training process. This, in turn, aids in minimizing the likelihood of such occurrences and guides the UAV base station's flight strategy.

Furthermore, $\xi_{\text{collision}}$ and ξ_{outage} are predefined hyperparameters that remain constant throughout the optimization process. Regarding interactions, the multi-agent reinforcement learning MASAC algorithm necessitates fitting the adjacent action-state value function. The reward function also relies on the communication signal-to-noise ratio and spectrum utilization efficiency of neighboring UAV base stations in its calculation process. To achieve this, it engages with Mj neighboring UAV base stations, including:

The coordinates of the UAV base station itself. The UAV base station's output actions. Two-dimensional relative positioning concerning the central user within the ground cluster. The signal-to-noise ratio of user information received by the cluster center. The spectral efficiency of the UAV base station at the current time. This section will introduce the Multi-Agent Soft Actor-Critic (MASAC) algorithm, a form of multi-agent maximum entropy reinforcement learning, based on the aforementioned trajectory regulation design using multi-agent reinforcement learning. Additionally, it will explore fusion ensemble learning and course learning techniques aimed at enhancing the training stability and convergence speed of the algorithm.

B. MAXIMUM ENTROPY REINFORCEMENT LEARNING FOR MULTIPLE AGENTS

In the face of a dynamic and uncertain emergency communication network environment, reinforcement learning leverages the Markov decision process to create a model. It acquires observations from the environment, which form the state s_t. Subsequently, it selects a strategy based on the action $\pi(a_t|s_t)$, produced by the policy π, to control the flight trajectory of the UAV base station. The agent then executes these actions, engages in interactions with the environment, assesses communication network coverage performance, and computes the reward function rt.

As the environment transitions from its current state s_t to the next state s_{t+1}, facilitated by the state transition distribution $(s_{t+1}|s_t, a_t)$ at time t, the action selection strategy of the reinforcement learning agent becomes closely tied

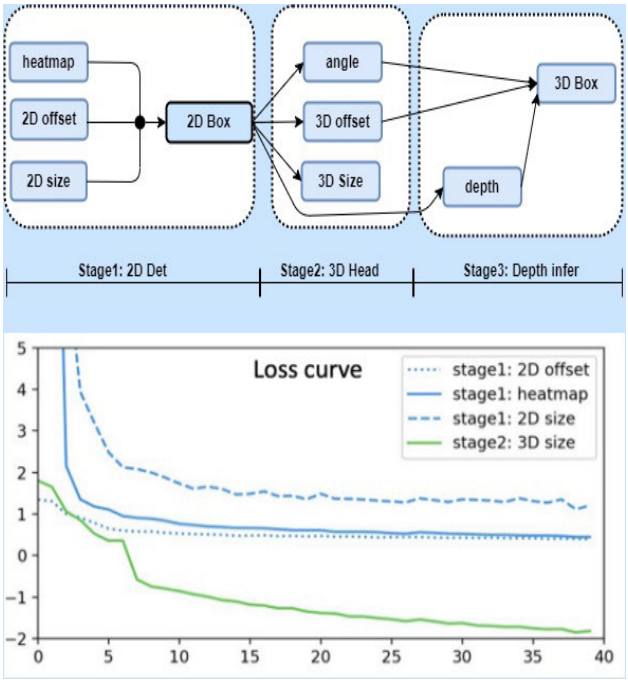

FIGURE 9. Accelerated convergence technical task division based on course learning.

to the state-action value function $Q(s_t, a_t)$. This function characterizes the expected cumulative reward over the long term when selecting action at for the UAV base station under state st, taking into account the extended period of emergency communication network coverage

$$Q(s_t, a_t) = r_t + \gamma E_{S_{t+1}\sim p_s}[V(s_{t+1})] \tag{30}$$

The function V(st) at time t represents the state value function, which serves as a metric to describe the anticipated value of long-term rewards for emergency communication network coverage performance that the UAV base station can achieve, starting from state st. The parameter γ denotes the discount factor, and it ensures the convergence of the reinforcement learning strategy iteration when it satisfies the condition $0 < \gamma \leq 1$. The state value function is essential for this convergence process.

$$V(s_t) = E_{a_t\sim\pi}[Q(s_t, a_t) - \alpha log\pi(a_t \mid s_t)] \tag{31}$$

The term $\alpha log\pi(a_t|s_t)$, at time t represents an entropy regularization component. This entropy regularization aligns with the optimization process of the action selection strategy. The algorithm's strategy output exhibits multi-modal characteristics, effectively addressing the dynamics and complexity of the learning environment, ultimately enhancing algorithm convergence stability. The parameter α in the entropy regularization term is the temperature factor, and its influence weight can be self-adjusted.

In scenarios with multiple agents within the network, agent i can solely access local observations o^t_i. The environmental state transition is influenced by the collective

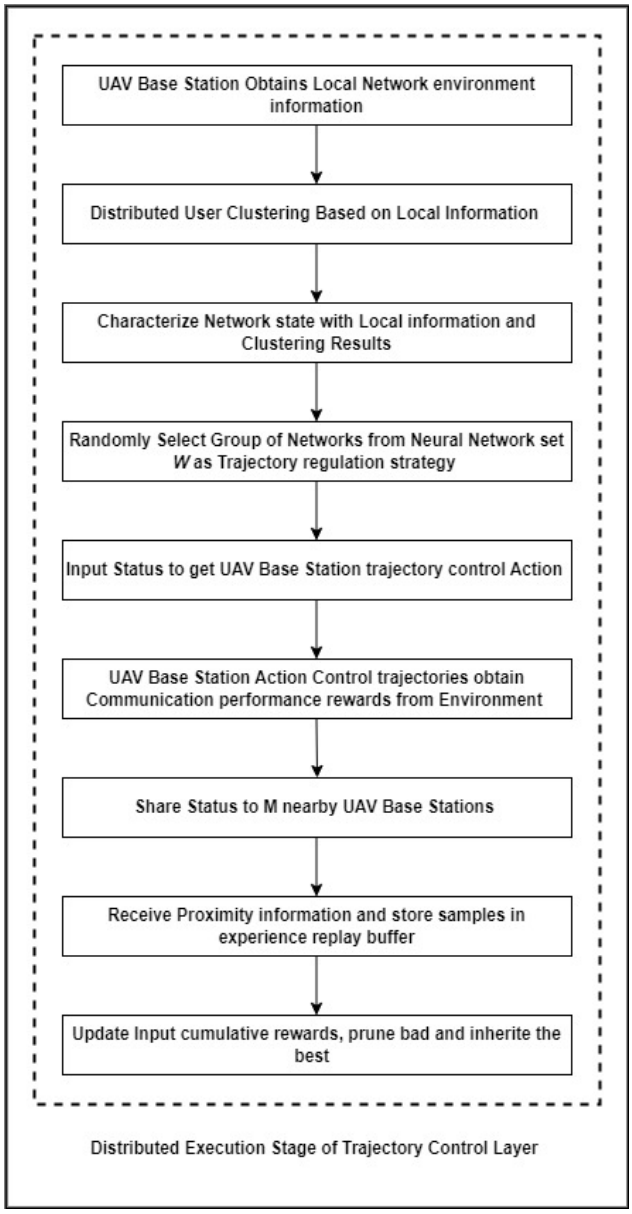

FIGURE 10. Overall process optimized for distributed coverage for large-scale post-disaster users.

action outputs of multiple agents simultaneously. Consequently, the environmental state transition distribution changes to agent i. In such a non-stationary state, conventional single-agent reinforcement learning struggles to converge. The multi-agent reinforcement learning MADDPG algorithm addresses this issue by fitting the global state-value function $Q(o_t^i, a_t^i, o_t^{-i}, a_t^{-i})$ using observations and output actions of other agents, stabilizing the agent i's learning environment. Here, -i represents all agents other than agent i.

This paper, building upon the Maximum Entropy Reinforcement Learning Soft Actor-Critic (SAC) algorithm and the Multi-Agent Deep Deterministic Policy Gradients (MADDPG) algorithm, adapts and enhances the adjacent state value function fitting from the SAC algorithm. This adaptation reduces communication overhead while ensuring algorithm convergence, facilitating distributed deployment of the algorithm.

Figure 5 depicts the architecture. All agents share the same Critic Network and Actor. In addition to this, we also keep the experiences of every agent in a common Replay Buffer. Every agent possesses an individual copy of its state data, observations from the environment, actions, and associated rewards. No other agent is aware of this information regarding a specific agent. But since the material in Replay Buffer is non-distinguishable, every agent gain from the collective experiences of all agents. Lastly, each agent updates the actor and critic networks asynchronously at each stage.

As depicted in Figure 6, each MASAC agent comprises six neural networks and one empirical replay buffer. The Actor network defines the action selection policy, denoted as $\pi\theta_{\mathrm{i}}$, with θ_{i} representing its neural network parameters. It takes the local observation state oti as input and generates the mean $(\mu\theta i)$ and standard deviation $(\sigma\theta_{\mathrm{i}})$ of the action output distribution for the observed state. This distribution is represented as $\pi\theta_{\mathrm{i}}$ and serves as the action selection strategy for agent i at time t.

The Double Q network consists of two neural networks: Critic1 and Critic2 networks, which estimate the state-value functions as $Q_{\theta_2^i}$ and $Q_{\theta_3^i}$, respectively. These neural networks have parameters θ_1^i and θ_3^i, respectively. By fitting two state-value functions, they mitigate the overestimation issue associated with a single Critic network, as discussed in reference [21] and [22].

The Target network encompasses three neural networks: Target Actor $\pi_{\theta_4^i}$, Target Critic1 $\pi_{\theta_5^i}$, and Target Critic2 $Q_{\theta_6^i}$. These networks share the same architecture as the actor network and the Critic networks but update their parameters at a slower rate. This deliberate slowing of parameter updates enhances training stability and accelerates algorithm convergence. The experience replay buffer is employed to store samples of agent interactions, where information about neighboring agents is obtained through inter-agent communication. During training, the agent samples from this replay buffer and randomly selects a sample set D to compute gradients for optimizing its objectives.

The objective of the action selection strategy is to maximize the state-action value function. Therefore, the optimization objective for the network can be formulated as follows

$$J_\pi\left(\theta_1^i\right) = E_{(o_t,a_t)\sim D}\left[\alpha \log \pi_{\theta_1}\left(\hat{a}_t^i \mid o_t^i\right) - \min_{i=2,3} Q_{\theta_i}\left(o_t^i, a_t^i, o_t^{-i}, a_t^{-i}\right)\right] \quad (32)$$

Given that the actor network generates a distribution function rather than a precise action value, it becomes necessary to represent the output action numerically when computing the gradient for the optimization objective. To accomplish this, employ the weighted parameter technique to derive an

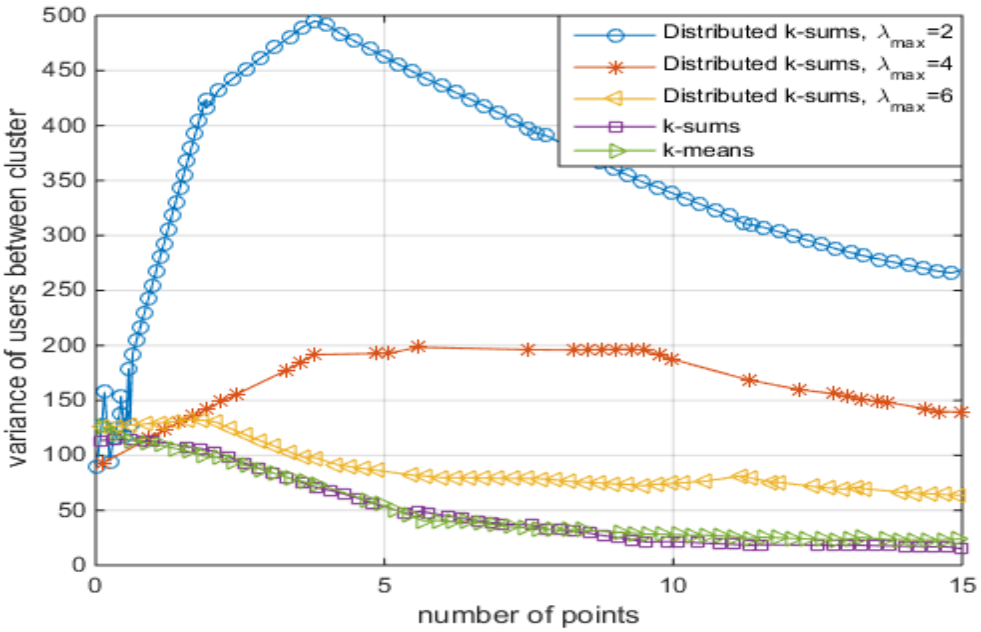

FIGURE 11. Effect of different clustering algorithms on the variance of the number of users between clusters.

estimated action value.

$$\hat{a}_t^i = \tanh\left(\mu_{\theta_i}\left(o_t^i\right) + \sigma_{\theta_i}\left(o_t^i\right)\epsilon_t^i\right) \tag{33}$$

, ε_t^i represents a Gaussian noise vector with a mean of 0, which is independent of the action selection strategy. The Critic network is designed to approximate the state-action value function, and thus, the optimization objective can be described in terms of the temporal difference error.

$$\begin{aligned} &J_Q\left(\theta_{2,3}^i\right) \\ &\quad = E_{(o_t,a_t,r_t,o_{t+1})\sim D}\Big[\Big(Q_{\theta_{2,3}^i}\left(o_t^i, a_t^i, o_t^{-i}, a_t^{-i}\right) \\ &\qquad - \Big(r_t + \gamma\Big(\min_{j=5.6} Q_{\hat{\theta}_{2,3}^i}\left(o_{t+1}^i, a_{t+1}^i, o_{t+1}^{-i}, a_{t+1}^{-i}\right) \\ &\qquad - \alpha\log\left(\hat{\pi}_{\theta_4^i}\left(a_{t+1}^i \mid o_{t+1}^i\right)\right)\Big)\Big)\Big)^2\Big]_{a_{t+1}\sim\hat{\pi}_{\theta_4^i}^i} \end{aligned} \tag{34}$$

Based on the above optimization objectives, the network parameters are updated.

Here, η represents the neural network update step. During the iterative exploration and training process, the agent acquires fresh samples from the environment and stores them in the experience replay buffer. Subsequently, it randomly selects batch samples from this buffer for training, following Equation (32) through Formula (34), enabling the agent to learn the optimal action selection strategy.

C. COMBINING LEARNING AND CURRICULUM LEARNING

The multi-agent reinforcement learning algorithm effectively addresses the non-stationary in multi-agent learning environments, and the MASAC algorithm adapts to complex dynamic settings. However, both multi-agent and maximum entropy reinforcement learning algorithms introduce complexity to neural networks. Therefore, this paper employs ensemble learning [23] and curriculum learning [24] techniques to enhance the speed and stability of algorithm convergence.

Stable Convergence Technique Based on Ensemble Learning: This approach combines ensemble learning, where multiple sets of neural networks are trained through bootstrapping. It collects feedback during the decision-making process, identifies sub-optimal networks for pruning, and retains high-performing networks to prevent catastrophic forgetting. This technique enhances the stability of the algorithm convergence process [26], [27]. Figure 7 provides a detailed overview of the implementation architecture for stable convergence technology based on ensemble learning. Each UAV base station's agents simultaneously train multiple sets of neural networks, forming an ensemble learning neural network set W.

$$\theta_1^i \leftarrow \theta_1^i + \eta_1\nabla_{\theta_1^i} J_\pi(\theta_1^i) \tag{35}$$

$$\theta_{2,3}^i \leftarrow \theta_{2,3}^i + \eta_{2,3}\nabla_{\theta_{2,3}^i} J_Q(\theta_{2,3}^i) \tag{36}$$

$$\theta_{4,5,6}^i \leftarrow \eta_{4,5,6}\theta_{4,5,6}^i + (1-\eta_{4,5,6})\theta_{1,2,3}^i \tag{37}$$

In the "distributed training" phase, independent sample sets from W groups, denoted as W1, W2, …, WD, are drawn from the experience replay buffer, and all neural networks within W undergo training. During the "distributed execution" phase, an agent randomly selects a neural network w from W to make decisions for the UAV base station, receive a reward rm, and update the cumulative reward w for the chosen

neural network w.

$$r_m^{(w)} = \tau_w r_m^{(w)} + (1 - \tau_w)\, r_m \tag{38}$$

Here, τ_w represents the update step for the cumulative reward of the neural network.

Furthermore, update the maximum cumulative reward r_m^{Wmax} within the neural network set W.

$$r_m^{W_{max}} = max\left(r_m^{W_{max}}, r_m^{(w)}\right) \tag{39}$$

If the cumulative reward $r_m^{(w)}$ of neural network w significantly lags behind the maximum cumulative reward ($r_m^{W_{max}}$) within the neural network set, pruning is initiated on neural network w. The neural network with the highest cumulative reward value among the remaining networks in W is then duplicated to replace the pruned neural network w. Through this ensemble learning design, MASAC agents can identify and prune neural networks that have suffered catastrophic forgetting, leading to significant performance degradation during training. This selection of neural network inheritance helps expedite the algorithm's convergence process.

Accelerated Convergence Technique Based on Curriculum Learning: This approach divides learning tasks into multiple sub-tasks, arranged from easy to difficult based on their physical significance. It designs reward functions for each sub-task, ranging from simple to complex, to reduce learning complexity and enhance algorithm convergence speed. Employing the concept of curriculum learning, as depicted in Figure 9, the reward function introduced in section IV is segmented into three sub-tasks: UAV base station's task of maintaining flight within a fixed area. The objective of reduce communication service interruptions by adjusting the UAV base station's flight trajectory, where interruptions occur when the signal-to-noise ratio at the UAV base station receiving user information from the cluster center falls below a threshold.

Optimizing flight trajectories of the UAV base station to maximize network spectral efficiency. Consequently, the reward functions for these three sub-tasks can be designed as follows:

$$r_A = -\xi_{collision} I^j_{collision} \tag{40}$$

$$r_B = -\xi_{collision} I^j_{collision} - N_j P_{outage}(t)\, \xi_{outage} I^j_{outage} \tag{41}$$

$$r_C = R_j(t) - \xi_{collision} I^j_{collision} - N_j P_{outage}(t)\, \xi_{outage} I^j_{outage} \tag{42}$$

It's important to highlight that learning more complex course material may lead neural networks to forget what they've learned in simpler lessons, potentially resulting in catastrophic forgetting. In the reward design for these advanced courses, it's essential to incorporate rewards from simpler courses, as demonstrated in Equation (38) and Equation (39). This collaborative approach, combined with the sub-network pruning technique of integrated learning, helps mitigate the impact of catastrophic forgetting.

The MASAC-based multi-UAV trajectory distributed regulation algorithm, which integrates ensemble learning and curriculum learning techniques, is presented in Algorithm 3. This algorithm effectively reduces the frequency of communication interruptions within the network, ultimately enhancing network spectral efficiency.

Algorithm 3 *MASAC-based Multi-UAV Trajectory Distributed Regulation Algorithm*

1. *Initialize the time step t to 0.*
2. *Loop while t is less than or equal to t:*
3. *Retrieve the observed state o_i^t from the environment.*
4. *Randomly select a set of neural networks from the ensemble learning neural network set W and input the observation state o_i^t into the actor-network to produce the action a_i^t.*
5. *Execute the selected actions, interact with the environment, and obtain the signal-to-noise ratio and spectral efficiency of user information at the cluster center at the current moment.*
6. *Communicate with neighboring drones, calculate the reward for the current curriculum learning task, and update the state $o_i^t + 1$.*
7. *Record samples and store them in the experience replay buffer.*
8. *Update the cumulative reward r_m^w and the maximum cumulative reward r_m^{Wmax} based on Equation (38) and Equation (39) to determine whether to proceed to the next curriculum learning stage.*
9. *If r_m^w is significantly lower than r_m^{Wmax}, perform pruning and inheritance operations on neural network w.*
10. *For each neural network in $W(1 \le n \le N)$, do:*
11. *Retrieve a batch of samples DN from the experience replay buffer.*
12. *Update the MASAC multi-agent reinforcement learning neural network parameters following Equation (35) to Equation (37).*
13. *Increment t by 1.*
14. *End the loop when t reaches T.*

D. COMPLEXITY ANALYSIS

During the "distributed execution" stage, each UAV base station must acquire its local state information and share it with neighboring UAV base stations. This complexity is directly related to the number of adjacent UAV base stations, denoted as M_j. Therefore, the algorithm's complexity at this stage can be expressed as $O(M_j)$. neural network needs to calculate gradients proportional to the batch sample size taken from the experience replay buffer, denoted as ND. Consequently, the complexity of the algorithm during this phase can be represented as O(WN). Given that the number of adjacent UAV base stations, M_j, is significantly smaller than the number of batch samples, ND, the overall complexity of Algorithm 3 can be described as O(WN). In the "distributed training"

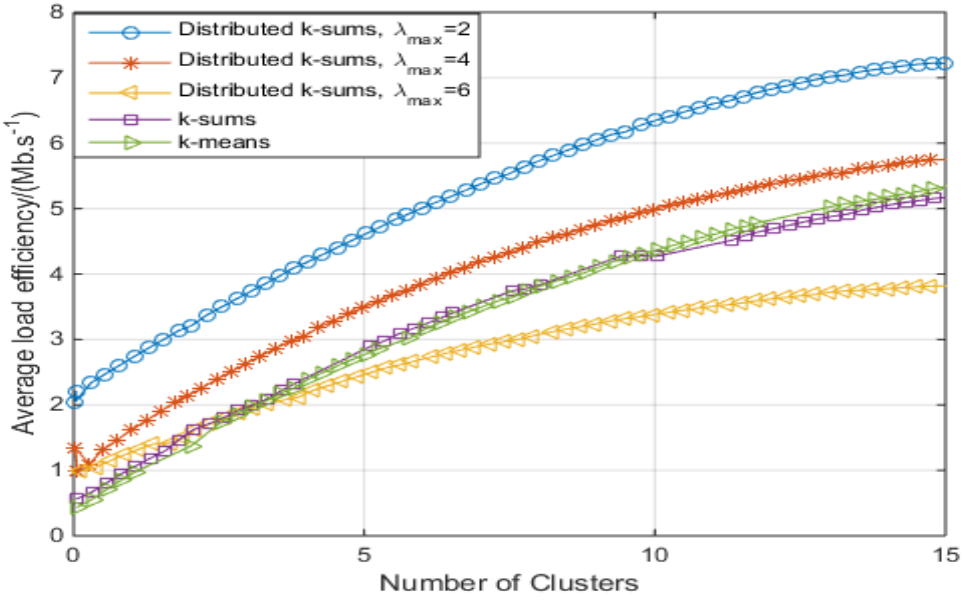

FIGURE 12. Convergence performance of MASAC algorithm averaging cumulative reward.

phase, each UAV base station is required to update all W neural networks within the ensemble learning neural network set W. The number of times each neural network needs to calculate gradients is proportional to the batch sample size taken from the experience replay buffer, denoted as ND. Consequently, the complexity of the algorithm during this phase can be represented as O(WN). Given that the number of adjacent UAV base stations, M_j, is significantly smaller than the number of batch samples, ND, the overall complexity of Algorithm 3 can be described as O(WN).

E. DISTRIBUTED COVERAGE OPTIMIZATION PROCESS FOR LARGE-SCALE POST-DISASTER USERS

The proposed distributed intent-based coverage optimization architecture in this paper can be categorized into two main layers: the network feature layer and the trajectory regulation layer. The network feature layer serves as the feature extraction stage for multi-agent reinforcement learning and is jointly realized through two algorithms: the user difference learning algorithm based on Bayesian inference (Algorithm 1) and the distributed k-sums algorithm considering user difference (Algorithm 2).

On the other hand, the trajectory regulation layer functions as the strategy implementation stage for the multi-agent reinforcement learning segment and is executed by the MASAC-based multi-UAV trajectory distributed regulation algorithm (Algorithm 3). The comprehensive workflow of distributed coverage optimization for large-scale post-disaster users is depicted in Figure 10.

V. SIMULATION ANALYSIS

This section evaluates the proposed aerial coverage architecture for large-scale post-disaster users, which is based on multi-agent reinforcement learning, and assesses the effectiveness of the corresponding algorithm through simulation experiments. We assume a scenario with 500 ground users located within a 1km × 1km area in the disaster-stricken region. The flight altitude range for the UAV base station is set between 100 m to 1,000 m. For the MASAC algorithm, both the Actor and Critic networks employ three fully connected layers in the hidden layer, with 512, 256, and 128 hidden neurons, respectively. The verification of the proposed large-scale post-disaster user-distributed coverage optimization scheme based on multi-agent reinforcement learning is conducted on the Python 3.7 platform. The Bayesian inference and distributed k-sums algorithm are implemented using the Numpy toolkit, while the MASAC algorithm for multi-agent reinforcement learning is implemented using the TensorFlow toolkit. The computing environment comprises Windows 10, an Intel 7th CPU, and a GTX 1060 GPU.

Firstly, validate the effectiveness of the distributed K-Sums clustering algorithm, which considers user differences in underlying optimization. Afterward conduct simulation experiments under varying maximum priority parameters $\lambda_{\max}$ and compare the results with the K-Sums algorithm and the K-Means algorithm. Figure 11 illustrates the impact of different clustering algorithms on the variance in the number of users between clusters. It is evident from Figure 14 that the proposed distributed K-Sums algorithm maintains a cluster balance similar to the K-Sums algorithm. When user information differences are not considered ($\lambda_{\max} = 1$), the variance in the number of users between clusters in the distributed K-Sums algorithm is nearly the same as that in the K-Sums algorithm, and considerably smaller than that in the K-Means algorithm. However, as the maximum priority parameter λ max increases, the proposed algorithm tends to

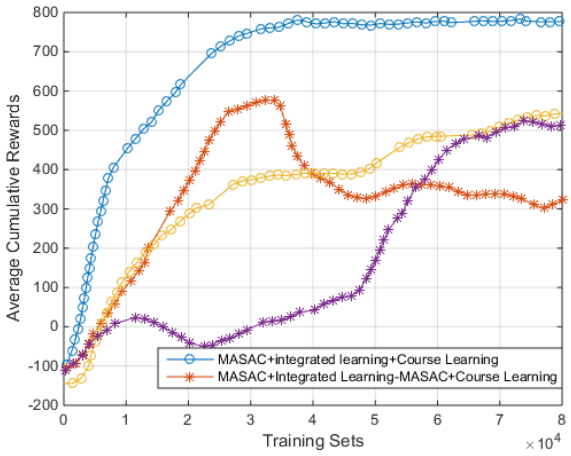

FIGURE 13. Convergence performance of MASAC algorithm averaging cumulative rewards.

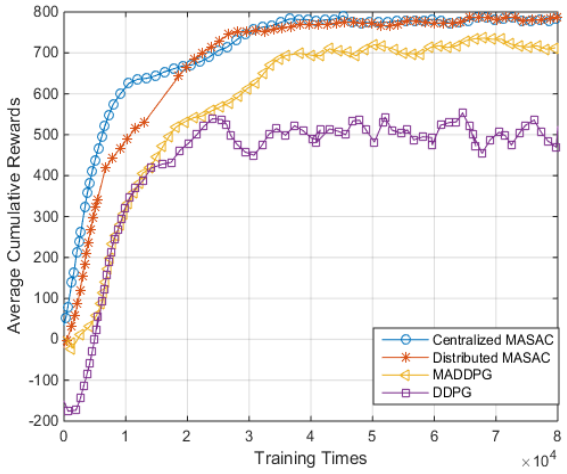

FIGURE 14. Convergence performance of average cumulative reward of different reinforcement learning algorithms.

prioritize users with higher priority parameters, potentially sacrificing some cluster balance. Consequently, the variance in the number of users between clusters increases.

Figure 15 illustrates the influence of different clustering algorithms on the average load efficiency of users within a cluster. As shown in Figure 15, as the number of clusters increases, the average intra-cluster distance decreases, leading to a significant increase in the average load efficiency for all clustering algorithms. When considering user information differences ($\lambda_{\max} = 1$), the average clustering efficiency of the proposed distributed K-Sums algorithm and the K-Sums algorithm is similar, and both outperform the K-Means algorithm overall. However, as the maximum priority coefficient $\lambda_{\max}$ increases, the algorithm utilizes Bayesian inference to learn the differences in user information. It assigns greater weight to users with higher priority coefficients when calculating dissimilarity measures, thus enhancing the average load efficiency. Based on these simulation results, this paper demonstrates that increasing the

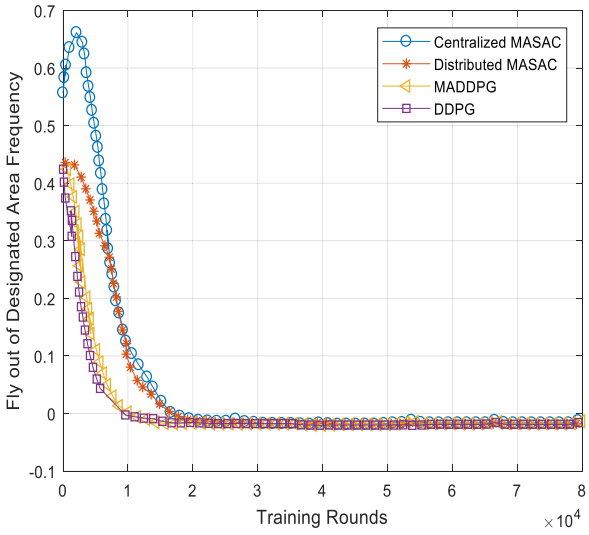

FIGURE 15. The learning effect of different reinforcement learning algorithms on the frequency of task 1– flying out of the specified region.

maximum priority coefficient λ max, it can enhance the communication efficiency for users with higher traffic demands and improve the overall average load efficiency within clusters. This verifies that the proposed algorithm effectively adapts to various priority services.

Additionally, simulate and verify the effectiveness of the upper-layer aerial coverage optimization algorithm based on multi-agent reinforcement learning as proposed in this paper. Figure 16 presents the convergence performance of the MASAC algorithm's average cumulative reward, showcasing the impact of ensemble learning and curriculum learning on MASAC's convergence rate and stability within the same simulation environment. The average cumulative reward serves as a crucial indicator for assessing the convergence of reinforcement learning algorithms [25]. It represents the average value of the reward function obtained across all time slots within a training round. Its specific physical interpretation depends on the design of the reward function. In this paper, the average cumulative reward reflects the sum of average spectral efficiency, average communication interruption penalties, and security penalties within a training round. As observed in Figure 12, both ensemble learning and curriculum learning contribute to increased algorithm convergence rates. However, ensemble learning directly tackles complex tasks and may converge to local optimal strategies with only average performance. Curriculum learning, on the other hand, exhibits catastrophic forgetting after learning Task 1 and Task 2, limiting further improvements in convergence performance. In contrast, MASAC algorithms that combine ensemble learning and curriculum learning can converge to superior strategies with faster convergence while mitigating the impact of catastrophic forgetting.

Figures 13 through 16 demonstrate the impact of various reinforcement learning algorithms on the trajectory regulation learning process of UAV base stations. The primary

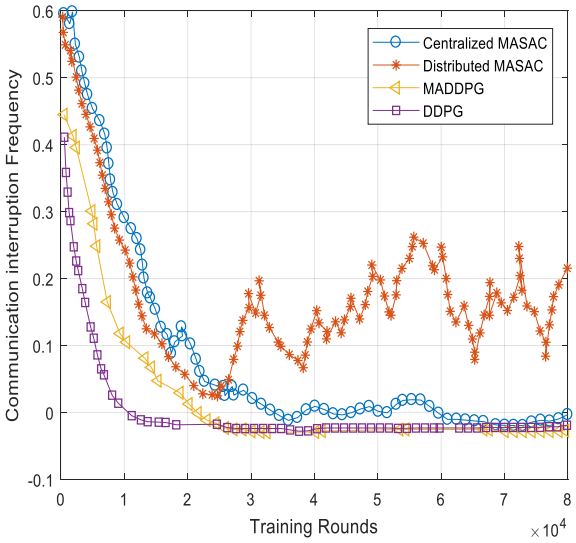

FIGURE 16. Learning effect of different reinforcement learning algorithms on task 2– communication interruption probability.

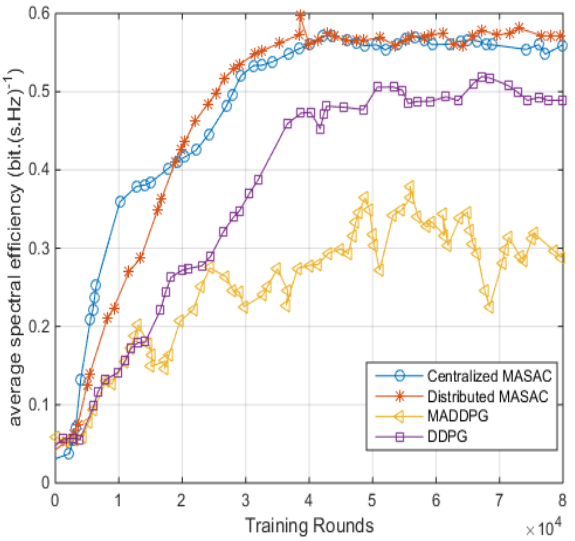

FIGURE 17. Learning effect of different reinforcement learning algorithms on task 3– average spectral efficiency.

comparison involves the proposed MASAC algorithm with the MADDPG algorithm [13] and the DDPG algorithm [11]. Figure 13 illustrates the convergence performance of the average cumulative reward for different reinforcement learning algorithms. Meanwhile, Figures 14 present changes in key indicators for course learning tasks 1 through 3, which encompass the frequency of UAV base station flights outside specified areas, communication interruption frequency, and average spectral efficiency.

Observing Figures 15 through 17, it becomes evident that the single-agent reinforcement learning DDPG algorithm rapidly accomplishes the learning for Task 1, which involves flying within the confined 1 km × 1 km area. However, it encounters difficulties in further learning Task 2 and Task 3. This is primarily because the strategic learning for the flight area of each UAV base station does not influence the flight areas of other UAV base stations, rendering the learning environment stable.

In contrast, for Task 2 and Task 3, the changes in UAV base station flight strategies affect the communications of other UAV base stations, creating an unstable learning environment. When comparing the multi-agent reinforcement learning MASAC algorithm and the MADDPG algorithm, both algorithms complete the learning for Task 1 and Task 2. However, the MADDPG algorithm exhibits poor convergence performance and stability due to its deterministic strategy algorithm. Additionally, the MADDPG algorithm's learning performance for Task 3, focusing on spectral efficiency, lags behind that of the MASAC algorithm. Furthermore, the simulation compares the centralized MASAC algorithm, which obtains the global state, with the distributed MASAC algorithm, which acquires the adjacent state. It is noteworthy that the distributed MASAC algorithm achieves a similar level of convergence as global optimization while substantially reducing communication overhead, as it only requires the status of neighboring drone base stations.

Figure 14 illustrates the impact of the number of UAV base stations on average spectral efficiency. From Figure 14, it becomes evident that the spectral efficiency of the DDPG and MADDPG algorithms decreases as the number of UAV base stations increases. This decline can be attributed to the increased complexity and non-stationary of the learning environment, which challenges the effectiveness of these algorithms. As the number of UAV base stations rises, both DDPG and MADDPG algorithms exhibit reduced spectral efficiency. Conversely, the MASAC algorithm proposed in this paper achieves higher spectral efficiency by jointly regulating the flight trajectories of UAV base stations when the number of UAV base stations is small. However, as the number of UAV base stations continues to increase, each UAV base station experiences interference from more neighboring UAV base stations, resulting in a decline in spectral efficiency. Furthermore, when comparing the centralized MASAC algorithm with the distributed MASAC algorithm, distributed optimization can achieve similar or even improved performance compared to global optimization. This advantage arises due to lower state input dimensions and smaller neural network sizes in scenarios with a large number of drones.

VI. CONCLUSION

This paper introduced a distributed intent-based aerial coverage optimization architecture aimed at facilitating the recovery of emergency communication for large-scale disaster-stricken users. The architecture consists of two layers: the network feature layer, responsible for user clustering using a distributed K-SUMS clustering algorithm tailored to account for user differences, and the trajectory control layer,

which optimizes the flight trajectories of UAV base stations using a distributed trajectory control algorithm based on multi-agent reinforcement learning, specifically the MASAC algorithm. Integrated learning and course learning techniques are integrated into MASAC to enhance convergence speed and effectiveness.

Simulation results demonstrate that the network feature layer algorithm can effectively handle user dynamics and differences, yielding clustering outcomes with improved average load efficiency. Additionally, the trajectory optimization layer algorithm designed in this paper can address the non-stationary learning environment for multiple UAV base stations. It optimizes the flight trajectories of each UAV base station based on proximity observations, reducing communication interruptions, enhancing spectral efficiency, and optimizing emergency network coverage performance.

While this research provides a distributed intent-based solution for restoring communication coverage to large-scale post-disaster users, there are still some limitations. Future research can explore the following two directions:

The proposed algorithm relies on multiple hyper-parameters, such as the number of adjacent UAV base stations, the number of observable users per UAV base station, and the correlation coefficient between UAVs. These hyper-parameter values are determined based on rules, but they could be further refined by introducing techniques such as attention mechanisms from deep learning.

This paper primarily focuses on user coverage optimization to swiftly re-establish communication in disaster areas. However, it does not consider other practical issues that may arise, including power constraints. Future research can delve into comprehensively addressing multiple optimization goals, building upon the foundation laid in this paper.

REFERENCES

[1] M. Matracia, N. Saeed, M. A. Kishk, and M.-S. Alouini, "Post-disaster communications: Enabling technologies, architectures, and open challenges," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1177–1205, 2022.

[2] A. Fascista, "Toward integrated large-scale environmental monitoring using WSN/UAV/crowdsensing: A review of applications, signal processing, and future perspectives," *Sensors*, vol. 22, no. 5, p. 1824, Feb. 2022.

[3] Y. Zhou, L. Liu, L. Wang, N. Hui, X. Cui, J. Wu, Y. Peng, Y. Qi, and C. Xing, "Service-aware 6G: An intelligent and open network based on the convergence of communication, computing and caching," *Digit. Commun. Netw.*, vol. 6, no. 3, pp. 253–260, Aug. 2020.

[4] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Shen, and C. Miao, "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 213–250, 1st Quart., 2023.

[5] Z. Ping, X. Xiao-Dong, H. Shu-Jun, N. Kai, X. Wen-Jun, and L. Yue-Heng, "Entropy reduced mobile networks empowering industrial applications," *J. Beijing Univ. Posts Telecommun.*, vol. 43, no. 6, pp. 1–9, 2020.

[6] Y. Zhou, L. Tian, L. Liu, and Y. Qi, "Fog computing enabled future mobile communication networks: A convergence of communication and computing," *IEEE Commun. Mag.*, vol. 57, no. 5, pp. 20–27, May 2019.

[7] Q. Chen, "Joint position and resource optimization for Multi-UAV-Aided relaying systems," *IEEE Access*, vol. 8, pp. 10403–10415, 2020.

[8] S. Yin, Y. Zhao, and L. Li, "Resource allocation and basestation placement in cellular networks with wireless powered UAVs," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1050–1055, Jan. 2019.

[9] T. Zhang, J. Lei, Y. Liu, C. Feng, and A. Nallanathan, "Trajectory optimization for UAV emergency communication with limited user equipment energy: A safe-DQN approach," *IEEE Trans. Green Commun. Netw.*, vol. 5, no. 3, pp. 1236–1247, Sep. 2021.

[10] K. Li, W. Ni, E. Tovar, and M. Guizani, "Deep reinforcement learning for real-time trajectory planning in UAV networks," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2020, pp. 958–963.

[11] W. Zhang, Q. Wang, X. Liu, Y. Liu, and Y. Chen, "Three-dimension trajectory design for multi-UAV wireless network with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 600–612, Jan. 2021.

[12] U. Challita, W. Saad, and C. Bettstetter, "Interference management for cellular-connected UAVs: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2125–2140, Apr. 2019.

[13] N. Zhao, Y. Cheng, Y. Pei, Y.-C. Liang, and D. Niyato, "Deep reinforcement learning for trajectory design and power allocation in UAV networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[14] H. V. Abeywickrama, Y. He, E. Dutkiewicz, B. A. Jayawickrama, and M. Mueck, "A reinforcement learning approach for fair user coverage using UAV mounted base stations under energy constraints," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 67–81, 2020.

[15] R. Lowe, "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[16] K. Sharma, B. Singh, E. Herman, R. Regine, S. S. Rajest, and V. P. Mishra, "Maximum information measure policies in reinforcement learning with deep energy-based model," in *Proc. Int. Conf. Comput. Intell. Knowl. Economy (ICCIKE)*, Mar. 2021, pp. 19–24.

[17] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[18] E. Dahlman, S. Parkvall, and J. Skold, *4G, LTE-Advanced Pro and The Road to 5G*. New York, NY, USA: Academic, 2016.

[19] C. Li, T. Trinh, L. Wang, C. Liu, M. Tomizuka, and W. Zhan, "Efficient game-theoretic planning with prediction heuristic for socially-compliant autonomous driving," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 10248–10255, Oct. 2022.

[20] Y. Wang, "McDPC: Multi-center density peak clustering," *Neural Comput. Appl.*, vol. 32, pp. 13465–13478, Feb. 2020.

[21] W. Shi, S. Song, and C. Wu, "Soft policy gradient method for maximum entropy deep reinforcement learning," 2019, *arXiv:1909.03198*.

[22] O. Peer, "Ensemble bootstrapping for Q-learning," in *Proc. Int. Conf. Mach. Learn. (PMLR)*, 2021, pp. 1–10.

[23] H. M. Gomes, J. P. Barddal, F. Enembreck, and A. Bifet, "A survey on ensemble learning for data stream classification," *ACM Comput. Surveys*, vol. 50, no. 2, pp. 1–36, Mar. 2018.

[24] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *J. Mach. Learn. Res.*, vol. 10, no. 7, pp. 1–53, 2009.

[25] P. R. Montague, R. S. Sutton, and A. G. Barto, "Reinforcement learning: An introduction," *Trends Cognit. Sci.*, vol. 3, no. 9, p. 360, 1999.

[26] Q. Hou and J. Dong, "Robust adaptive event-triggered fault-tolerant consensus control of multiagent systems with a positive minimum interevent time," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 7, pp. 4003– 4014, Jul. 2023.

[27] Q. Hou and J. Dong, "Distributed dynamic event-triggered consensus control for multiagent systems with guaranteed performance and positive inter-event times," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 1, pp. 746–757, Jan. 2024.

RAHUL SHARMA received the B.E. degree in electronics and communication engineering from Rajiv Gandhi Technical University, Bhopal, India, in 2006, and the M.Tech. degree from the ABV-Indian Institute of Information Technology and Management, Gwalior, India, in 2009. He is currently an Assistant Professor with Lovely Professional University, Phagwara, India. His main research interest includes wireless networks and optimization.

SHAKTI RAJ CHOPRA received the M.B.A. degree in IT and marketing from Agra University, the M.Tech. degree in control and instrumentation from NIT Allahabad, and the Ph.D. degree in electronics and electrical engineering from Lovely Professional University, Phagwara. He worked at different levels, including the Head of the Department of Wireless Communication, School of Electronics and Electrical Engineering, Lovely Professional University. He is currently a Professor with Lovely Professional University, with more than 18 years of experience in academics. He has taught various courses to UG and PG levels, such as advanced wireless communication systems, electromagnetic field theory, microwave and antenna, and digital electronics. He has published more than 45 research papers in refereed IEEE, Springer, and IOP Science journals and conferences. He has supervised more than 18 master's thesis and more than 50 bachelor's student projects. He has taken and completed ten non-government and consultancy projects. He has attended/participated in 24 national/international online webinars. His research interests include massive MIMO, cognitive radio, blockchain, AI, and machine learning. He is a Reviewer of most reputed journals, such as *Applied Soft Computing*, IEEE Access, and *China Communication*. He has organized several workshops, summer internships, and expert lectures for students.

AKHIL GUPTA (Senior Member, IEEE) received the B.E. degree in electronics and communication engineering from Jammu University, Jammu and Kashmir, India, in 2010, the M.Tech. degree in electronics and communication engineering from the Jaypee University of Information Technology, Waknaghat, India, in 2013, and the Ph.D. degree in electronics and communication engineering from Shri Mata Vaishno Devi University, Jammu and Kashmir, in 2017.

He is currently an Associate Professor with the School of Electronics and Electrical Engineering, Lovely Professional University, Punjab, India. He has guided three Ph.D. students and many bachelor's and master's students. He has published more than 50 research papers, including IEEE Transactions, IEEE journals, and international conference papers. His research interests include the physical and network layer perspective of 5G and 6G communications technologies, such as NOMA, cooperative wireless communications, device-to-device communications, unmanned aerial vehicles, massive MIMO, cognitive radio, and ultra-reliable low latency communication. He is also working on the security issues of next-generation networks.

Dr. Gupta is a member of the International Association of Engineers and the Universal Association of Computer and Electronics Engineers. He received the Teaching Assistantship from the Ministry of Human Resource Development, from 2011 to 2013. He is the author of the most popular paper in IEEE Xplore and an active reviewer of many IEEE, Springer, Elsevier, and Wiley journals. He has also served as a keynote speaker, the session chair, and a technical program committee member for various international conferences. He has more than 3400 citations in his credit. He has been listed in the Top 2% scientist of the world by Stanford University in three consecutive years, from 2020 to 2023.

RUPENDEEP KAUR received the B.Tech. degree in electronics and communication engineering from the Beant College of Engineering and Technology, Gurdaspur, India, in 2008, and the M.Tech. degree from Guru Nanak Dev University, Amritsar, India, in 2010. She is currently an Assistant Professor with Guru Nanak Dev University. Her main research interests include optical communication and wireless communication.

SUDEEP TANWAR (Senior Member, IEEE) received the B.Tech. degree from Kurukshetra University, India, in 2002, the M.Tech. degree (Hons.) from Guru Gobind Singh Indraprastha University, Delhi, India, in 2009, and the Ph.D. degree in wireless sensor network, in 2016. He is currently a Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University, India. He is also a Visiting Professor with Jan Wyzykowski University, Polkowice, Poland, and the University of Pitesti, Pitesti, Romania. He has authored two books, edited 13 books, and more than 450 technical papers, including top journals and top conferences, such as IEEE Transactions on Network Science and Engineering, IEEE Transactions on Vehicular Technology, IEEE Transactions on Industrial Informatics, IEEE Wireless Communications, *IEEE Network*, ICC, GLOBECOM, and INFOCOM. He initiated the research field of blockchain technology adoption in various verticals, in 2017. His H-index is 74. He actively serves his research communities in various roles. His research interests include blockchain technology, wireless sensor networks, fog computing, smart grids, and the IoT. He is a member of the Technical Committee on Tactile Internet of the IEEE Communication Society. He is a Senior Member of CSI, IAENG, ISTE, and CSTA. He received the Best Research Paper Awards from IEEE GLOBECOM 2018, IEEE ICC 2019, and Springer ICRIC-2019. He has served many international conferences as a member of the organizing committee, such as the Publication Chair for FTNCT-2020, ICCIC 2020, and WiMob2019; a member of the Advisory Board for ICACCT-2021 and ICACI 2020; the Workshop Co-Chair for CIS 2021; and the General Chair for IC4S 2019 and 2020 and ICCSDF 2020. He is a Final Voting Member of the IEEE ComSoc Tactile Internet Committee, in 2020. He is also serving on the editorial boards of *Computer Communications*, *International Journal of Communication System*, and *Security and Privacy*. He is leading the ST Research Laboratory, where group members are working on the latest cutting-edge technologies.

GIOVANNI PAU (Senior Member, IEEE) received the bachelor's degree in telematic engineering from the University of Catania, Italy, and the master's (cum laude) and Ph.D. degrees in telematic engineering from the Kore University of Enna, Italy. He is currently an Associate Professor with the Faculty of Engineering and Architecture, Kore University of Enna. He is the author/coauthor of more than 100 refereed papers published in journals and conference proceedings. His research interests include wireless sensor networks, fuzzy logic controllers, intelligent transportation systems, the Internet of Things, smart homes, and network security. He is a member of the IEEE (Italy Section) and has been involved in several international conferences as the session co-chair and a technical program committee member. He serves/served as a leading guest editor for the special issues of several international journals. He is an Editorial Board Member and an Associate Editor of several journals, such as IEEE Access, *Wireless Networks* (Springer), *EURASIP Journal on Wireless Communications and Networking* (Springer), *Wireless Communications and Mobile Computing* (Hindawi), and *Sensors* (MDPI).

GULSHAN SHARMA received the B.Tech., M.Tech., and Ph.D. degrees. He is currently a Senior Lecturer with the Department of Electrical Engineering Technology, University of Johannesburg. He is also a Y-Rated Researcher with NRF, South Africa. His research interests include power system operation and control and the application of AI techniques to power systems. He is an Academic Editor of *International Transactions on Electrical Energy System* (Wiley) and a Regional Editor of *Recent Advances in Electrical and Electronic Engineering* (Bentham Science).

FAYEZ ALQAHTANI is currently a Full Professor with the Department of Software Engineering, College of Computer and Information Sciences, King Saud University (KSU). He was appointed as the Director of the Computer Division, Deanship of Student Affairs. He is also a member of a number of academic and professional associations, such as the Association for Computing Machinery (ACM), the Australian Computer Society, and the Association for Information Systems. He has conducted research projects in several areas of information and communication technology, such as Web 2.0, information security, enterprise architecture, software process improvement, the Internet of Things, and fog computing. He has participated in several academic events.

AMR TOLBA (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Egypt, in 2002 and 2006, respectively. He is currently a full professor of computer science at King Saud University (KSU), Saudi Arabia. He has authored/coauthored over 200 scientific papers in top-ranked (ISI) international journals (such as IEEE IoT, ACM TOIT, IEEE CEMAG, IEEE ACCESS, IEEE SYSTEMS, FGCS, JNCA, NC&A, JAIHC, COMNET, COMCOM, P2PNET, VCOM, WWWJ). He served as a technical program committee (TPC) member at several conferences (Such as DSIT 2022, CICA2022, EAI MobiHealth 2021, DSS 2021, AEMCSE 2021, ICBDM 2021, ICISE 2021, DSS 2020, NCO 2020, ICISE2019, ICCSEA 2019, DSS 2019, FCES 19, ICISE 2018, ESG'18, Smart Data'17, NECO 2017, NC'17, WEMNET'17, NET'17, Smart Data'16). He has been included in the list of the top 2% of influential researchers globally (prepared by scientists from Stanford University, USA) during the calendar years 2020, 2021, 2022, and 2023, respectively. He has translated four books into the Arabic language. His main research interests include artificial intelligence (AI), the Internet of Things (IoT), data science, and cloud computing. He served as Associate Editor/Guest Editor for several ISI journals.

Open Access funding provided by 'Università degli Studi di Enna "KORE"' within the CRUI CARE Agreement