WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n3.06
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Joint Optimization of Resource Allocation and Trajectory in UAV-assisted Wireless Semantic communication System

Hanxiao Sun, Ping Xie

School of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

ABSTRACT

In this paper, we investigate the resource allocation design of an unmanned aerial vehicle (UAV)-enabled semantic communication system, in which UAVs are assigned to transmit semantic information to multiple user nodes. Our goal is to maximize the semantic information sum rate of the communication system by jointly optimizing the subcarrier allocation strategy and the trajectory of the UAVs, taking into account the minimum required semantic data rate for each user node, the minimum semantic similarity that can be received, the maximum cruise speed of the UAVs, and the initial/final position of the UAVs. The design is formulated as a generally tricky mixed-integer nonconvex optimization problem. Subsequently, a computationally efficient iterative algorithm is proposed to obtain a locally optimal solution, and further proofs demonstrate that the proposed algorithm is guaranteed to converge to a solution that satisfies at least the KKT condition of the original optimization problem.

KEYWORDS

Semantic communication; Unmanned aerial vehicle; Resource allocation; Trajectory optimization

1. INTRODUCTION

With the rapid development of communication technologies from first-generation (1G) to fifth-generation (5G), the achievable transmission rate has increased by tens of thousands of times, effectively addressing most of the challenges in 5G networks through emerging techniques such as higher millimeter wave (mmWave) frequencies [1], intelligent reflecting surfaces [2], edge Artificial Intelligence (AI) [3], and integrated terrestrial, airborne, and satellite networks [4]. However, in recent years, a large number of 6G applications with numerous nodes have emerged, such as collaborative robots, ultra-intelligent Internet of Things and metaverse [5-7]. The interconnection of these intelligent programs generates a large amount of data traffic (at the zetta-bytes scale), which increases the demand for bandwidth and requires faster system response and reliable, more efficient information exchange. However, a series of complex communication technologies that have been proposed are gradually approaching the Shannon limit, and cannot support many intelligent application scenarios in 6G mobile communication which poses new challenges for traditional wireless communication. Therefore, new communication technologies need to be developed.

Fortunately, semantic communication has received extensive attention in recent years [8]. It can integrate users, application requirements, and the meaning of information into data processing and transmission, and is considered a promising technology in the sixth-generation (6G) wireless networks [9]. Therefore, to cope with the growing data rate, reduce network communication costs, and meet the more stringent requirements for low-latency and high-reliability data transmission, a novel paradigm known as semantic communication is inspired as a brand new technology in 6G to break out of the Shannon's trap. Different from bit-based wireless communication systems, semantic communication further compresses data while retaining meaning by extracting the meaning of the data and filtering out content irrelevant to the predefined transmitted information through advanced artificial intelligence technologies [10-12], thus greatly reducing the amount of transmitted data while retaining the original semantics. At the same time, the semantic decoder deployed in the destination device can recover the true meaning from the received signal, even if there may be intolerable bit errors at the syntax level [13]. Correspondingly, semantic communication can significantly save the required bandwidth, improve resource allocation effectiveness, ensure sufficient communication reliability [14], and reduce the burden of wireless information data transmission in 6G networks, making it an outstanding participant in promoting the great leap forward of 6G networks.

Due to the evolution of deep learning (DL) techniques in understanding language, audio, and images, DL-based architectures have recently been developed. In particular, to improve the transmission reliability of different types of sources, a joint source and channel coding (JSCC) text transmission method is proposed, where an encoder and decoder based on deep learning (DL) were developed to handle semantic information contained in sentences [15]. The authors of [16] further proposed an advanced DL-based semantic communication tool (DeepSC) for text transmission, which outperformed traditional transmission methods at low and moderate signal-to-noise ratios (SNR). The authors of [17] developed a DL-based JSCC method for image transmission, where DL was used to encode image pixel values into transmitted symbols. The results showed that this method was superior to traditional image compression and coding methods at low SNRs. However, the research work of [15-17] used neural networks to simulate wireless channels and evaluated the performance of information transmission based on cross-entropy and mutual information in the loss function continuously trained, which cannot provide mathematical theoretical analysis for system design. Therefore, the authors of [18] defined a new performance metric based on [16], semantic rate, to characterize the transmission efficiency of semantic communication. Moreover, a semantic-aware resource allocation scheme is proposed, which maximize the spectral efficiency of semantic communication by optimizing the channel assignment and the number of transmitted semantic symbols [18]. Although many researchers pay attention to DL-enabled and utilize its potential in enhancing communication efficiency, semantic communication is still facing some challenges. For example, most existing semantic communication architectures are focused on short-range semantic information transmission between fixed nodes, and simulations often assume ideal channel state information (CSI) and rarely consider the significant path loss that semantic information will experience during long-distance propagation. This can cause significant performance degradation in actual semantic communication systems, hindering their end-to-end semantic information transmission efficiency.

In recent years, the rapid development of intelligent devices, sensors, and communication technology has promoted the explosive development of unmanned aerial vehicle (UAV) assisted applications, where low-altitude UAVs integrated with miniaturized communication transceivers can be used as aerial mobile base stations (BSs) [19-21] or relays [22,23] to improve the performance of terrestrial wireless communication systems. It can dynamically adjust its position and significantly shorten the communication distance between the UAV and ground users with appropriate UAV trajectory design, maintain line-of-sight (LoS) link connection with ground users, thereby greatly improve the link quality. Therefore, to support highly reliable semantic information transmission for future 6G systems, this paper proposes a UAV-enabled wireless semantic communication system. In fact, the semantic communication system supported by UAVs has also been partially studied. An efficient task-oriented semantic communication framework is designed at the semantic level, as well as a triplet-based image information scene graph [24]. A new personalized semantic encoder is designed based on user interest to meet the requirements of personalization salience [17]. It developed a new aerial image transmission paradigm for scene classification tasks. A lightweight model is established on the front-end UAV for semantic block transmission of perceived images and channel conditions. And the author exploit deep reinforcement learning to explore semantic blocks that best contribute to the backend classifier under different channel conditions. Nevertheless, a semantic communication system is usually evaluated in terms of both accuracy and efficiency. The above work aims to maximize task completion performance, while ignoring the transmission efficiency of the overall semantic communication system. In addition, a large number of smart application devices have high requirements for low latency. Therefore, how to improve the transmission efficiency while guaranteeing the above requirements is still an open question.

Although [25] has defined the system throughput for text transmission and formulated user affinity and spectrum allocation issues to maximize the defined system throughput, they have optimized resource allocation at the message level rather than the semantic level. To study resource allocation at the semantic level, the previous work developed an effective network resource allocation scheme [26-28]. However, due to the mobility of the drone, the distance between the drone and the users affects the signal-to-noise ratio received by the user, thereby affecting the semantic similarity. Therefore, the existing resource allocation scheme is not suitable for UAV-assisted wireless semantic communication systems. Besides, the current optimization of semantic communication systems usually uses methods such as optimizing neural network structure and parameters, or using pre-trained models that can better complete specific tasks [29,30], while ignoring the optimization of systems using traditional communication techniques and facilities.

In order to improve the transmission efficiency of the overall semantic communication system under the premise of satisfying the semantic accuracy and low latency of ground users, we study the resource allocation and trajectory optimization scheme of the UAV-assisted wireless semantic communication system, in which the semantic rate [18] is used as a performance index to evaluate the transmission efficiency of the overall system. Semantic rate is affected by bandwidth and semantic similarity which relies on the neural network structure of DeepSC and channel conditions. Therefore, this paper jointly optimizes the trajectories of UAV and resource blocks to maximize the overall network throughput while meeting the needs of ground users with low latency and high accuracy. The main contributions of this paper are organized as follows:

We study a UAV-enabled wireless semantic communication system, where UAV is deployed as a BS to directly transmit semantic information to all ground users. We use the curve fitting method to obtain the closed-form expression of semantic similarity. Based on the function, a joint UAV trajectory and resource allocation problem is formulated, whose goal is to maximize the throughput of the overall semantic communication system.

To solve the problem, we propose a computationally efficient iterative alternating algorithm to achieve a suboptimal solution. we first separate it into two sub-problems. The first sub-problem is a resource allocation optimization problem with a given UAV trajectory.

2. MODEL SYSTEM AND PROBLEM FORMULATION

2.1. System Model

As shown in Fig. 1, we consider a UAV-based downlink wireless semantic communication network, which consists of one rotary-wing UAV and a set of users denoted by $k \in \mathrm{K} \triangleq \{1, \cdots, K\}$. The UAV using semantic communication techniques transmit their text data to all users. Here, semantic communication techniques enable the UAV to transmit the meaning of their data to the users which will recover the original data according to its received data. The DeepSC transceiver is assumed to be trained at UAV. Then the trained semantic transmitter model is broadcast to users.

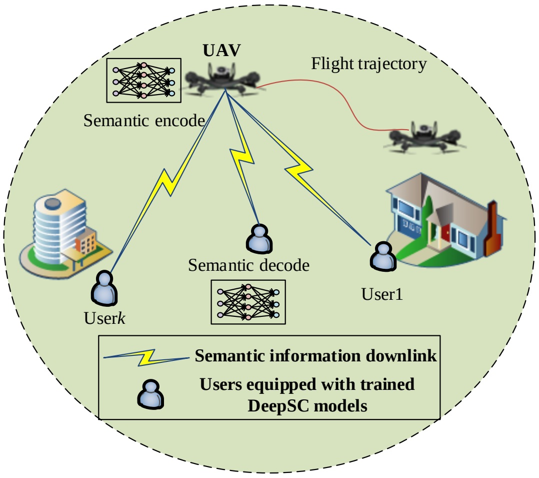

Fig. 1 UAV-assisted downlink semantic communication system model

2.1.1. Semantic communication model

In our model, the k-th user generates a sentence $\mathbf{s}_k = [w_{k,1}, w_{k,2}, \ldots, w_{k,l}, \ldots, w_{k,L_n}]$, where $w_{k,l}$ denotes the l-th word and Ln is the sentence length at the n-th user. Then the sentence is fed into the DeepSC transmitter and mapped to a semantic symbol vector $\mathbf{X}_k = [\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \ldots, \mathbf{x}_{k,o_k L_k}]$, where $\mathbf{X}_k \in \mathbb{R}^{o_k L_k \times 2}$ and $o_k L_k$ is the length of the semantic symbol vector for a sentence at the n-th user. We notice that the length of $\mathbf{X}_k$ varies with Ln to extract the semantic information of sentences with different lengths more effectively. In such a model, ok denotes the average number of semantic symbols used for each word at the k-th user, and each semantic symbol can be transmitted over transmission medium directly.

2.1.2. Channel model

An orthogonal frequency division multiple access (OFDMA) technique is adopted at the UAV to transmit semantic information. We assume that the system bandwidth is divided into I_F orthogonal subcarriers and each subcarrier can be allocated to at most one user to avoid multiple access interference. For mathematical description, we build a three-dimensional (3D) Cartesian coordinate express the locations of all nodes. We consider a finite time period T to guarantee the efficiency of semantic information transmission. The horizontal coordinate of each ground user is known in advance and fixed at $w_k = [x_k, y_k]$. In addition, it is also assumed that the UAV flies at a fixed altitude H meters and the maximum flying speed v_{max} in meters per second (m/s). The time-varying horizontal coordinate of the UAV over time instant t is denoted by $q(t) = [x(t), y(t)]$. However, the continuous variable t essentially implies an infinite number of UAV speed constraints which are difficult to solve generally. For the ease of design, the time T is divided into N time slots with equal-length, $\delta = T/N$. Each time slot is small enough to ensure the UAV position is approximately unchanged when flying at the maximum speed v_{max}. Thus, the 3D trajectory of UAV can be approximated by $q[n] = [x[n], y[n], H], n \in \{1, \cdots, N\}$. We consider the initial and final position of the UAV can be denoted by $\{q[I], q[F]\}$.

$$\boldsymbol{q}_I = \boldsymbol{q}[1], \tag{1}$$

$$\boldsymbol{q}_F = \boldsymbol{q}[N], \tag{2}$$

$$\|\boldsymbol{q}[n+1] - \boldsymbol{q}[n]\|^2 \le (S_{max})^2, n = 1, \ldots, N-1, \tag{3}$$

Where $S_{max} = v_{max}\delta$ is the maximum aviation distance of UAV within one time slot. Furthermore, the distance between the UAV and user k in time slot n is denoted by $d_k[n], \forall n, k$, which can be expressed as

$$d_k^2[n] = \| \boldsymbol{q}[n] - \boldsymbol{w}_k \|^2 + H^2, \forall n, k \ , \tag{4}$$

For simplicity, we assume that the communication links from the UAV to the ground users are dominated by the LoS links where the channel quality depends only on the UAV-user distance. Furthermore, the Doppler effect caused by the UAV mobility is assumed to be well compensated at the receivers. Thus, the channel power gain from UAV to user k during slot n follows the free-space path loss model, which can be expressed as

$$\boldsymbol{h}_k[n] = \sqrt{\frac{\beta_0}{\left(d_k[n]\right)^2}}, \forall n \ , \tag{5}$$

where β_0 represents the channel power gain at the reference distance $d_0 = 1$ m. Therefore, the receiving SNR of user k at the time slot n can be expressed as

$$\gamma_k[n] = \frac{P \mid h_k[n] \mid^2}{\sigma^2}, \forall n, k \ , \tag{6}$$

where σ^2 denotes the additive white Gaussian noise (AWGN) power at ground user nodes. Also, to reduce the peak-to-average power ratio in the considered multicarrier systems. In this work, we assume a fixed power allocation to simplify the resource allocation design. Adaptive power allocation will be considered in the future work. Therefore, we set the transmit power on each subcarrier at the UAV is P.

Denote $b_k^i[n] \in \{0,1\}$ as the binary subcarrier allocation variable of the i-th, $i \in \{1, ..., I_F\}$, subcarrier for user k at time slot n. We have $b_k^i[n] = 1$ if user k is allocated to subcarrier i, at time slot n and, $b_k^i[n] = 0$, otherwise. Furthermore, each subcarrier can be allocated to at most one user to avoid multiple access interference. Thus, in order to characterize the performance of semantic text transmission, the achieved semantic rate $S_k[n]$(susts/s/Hz) from the UAV to user k at time slot n over all subcarriers can be defined as

$$S_k[n] = \sum_{i=1}^{I_F} b_k^i[n] \frac{I}{OL} \varepsilon\left(O, \gamma_k[n]\right), \forall n, k \ , \tag{7}$$

Where L represents the expected number of words to be transmitted in each sentence, O denotes the average number of mapped semantic symbols for each word, I denotes the average amount of semantic information which the transmitted sentence contains, and $\varepsilon(O, \gamma_k[n])$ denotes the semantic similarity.

2.2. Problem Formulation

In this section, we first introduce the semantic rate as an important performance metric, and formulate the resource allocation problem for maximize the sum semantic rate of all users in UAV-based downlink wireless semantic communication network. we have

$$S_{sum}(B, \varepsilon) = \sum_{n=1}^{N} \sum_{k=1}^{K} S_k[n], \forall n, k \ , \tag{8}$$

It is observed that the sum semantic rate is effected on the number of subcarrier allocated to each user and the semantic similarity.

However, the objective function is intractable since there is a lack of an explicit form for the semantic similarity. To overcome the difficulty of lacking a closed-form expression, the data regression method was proposed in [31].

$$\varepsilon\left(O,\gamma\right) \approx \tilde{\varepsilon}_O\left(\gamma\right) \triangleq A_{O,1} + \frac{A_{O,2} - A_{O,1}}{1+e^{-\left(C_{O,1}\gamma + C_{O,2}\right)}}\ , \tag{9}$$

where $A_{O,1}, A_{O,2} > 0$ denote the lower asymptote and the upper asymptote respectively, and $C_{O,1} > 0$ denotes the logistic growth rate, and $C_{O,2}$ controls the logistic mid-point.

Based on equation (9), (8) can be transformed into

$$\tilde{S}_k[n] = \sum_{k=1}^{K} b_k^i[n] \frac{I}{OL}\left(A_1 + \frac{A_2 - A_1}{1+e^{-\left(C_{O,1}\gamma_k[n] + C_{O,2}\right)}}\right)\ , \tag{10}$$

where $\gamma_k[n] = \dfrac{\gamma_0 P}{\| \mathbf{q}[n] - \mathbf{w}_k \|^2 + H^2}$ and $\gamma_0 = \beta_0 / \sigma^2$.

$$\tilde{S}_{sum}(B,Q) = \sum_{n=1}^{N}\sum_{k=1}^{K} \tilde{S}_k[n], \forall n,k\ , \tag{11}$$

where $B = \{b_k^i[n], \forall i \in I_{\mathrm{F}}, k \in K, n \in N\}$. Obviously, the semantic similarity is effected by the UAV trajectory since the SNR changes following the different trajectory.

Due to the unstability of UAV-ground channel condition, a minimum limit of semantic rate is required to guarantee QoS in UAV-based wireless semantic communication networks. Thus, we should design the fly trajectory of UAV and the subcarrier allocation to meet the following condition.

$$\tilde{S}_k[n] \geq S_{th}^k, \forall n,k\ , \tag{12}$$

where S_{th} denotes the minimum required semantic rate of user k.

However, although satisfying (12) can make the semantic information successfully be transmitted to user k, it cannot guarantee that the user can succeed to recover the initial sentence from the received semantic symbol thereby affecting the accuracy of the received sematic information. Thus, we assume all user satisfy the following condition.

$$\tilde{\varepsilon}_O\left(\gamma\right) \geq \varepsilon_{th}^k, \forall n,k\ , \tag{13}$$

where ε_{th}^k represent the minimum required semantic similarity.

Thus, we maximize the sum semantic rate via jointly optimizing the UAV trajectory and subcarrier allocation design. The optimization problem can be expressed as follows:

$$\text{P0} \qquad \max_{\boldsymbol{Q},\boldsymbol{B}}\ \tilde{S}_{sum}(B,Q)\ , \tag{14a}$$

$$s.t. \qquad \sum_{n=1}^{N} b_k^i[n] \frac{I}{OL} \tilde{\varepsilon}_O\left(\gamma\right) \geq S_{th}^k, \forall n,k, \tag{14b}$$

$$\tilde{\varepsilon}_O\left(\gamma\right) \geq \varepsilon_{th}^k, \forall k\ , \tag{14c}$$

$$\sum_{k=1}^{K} b_k^i[n] \le 1, \forall n, i, \tag{14d}$$

$$b_k^i[n] \in \{0,1\}, \forall n, k, i, \tag{14e}$$

$$\mathbf{q}_I = \mathbf{q}[1], \tag{14f}$$

$$\mathbf{q}_F = \mathbf{q}[N], \tag{14g}$$

$$\| \mathbf{q}[n+1] - \mathbf{q}[n] \|^2 \le \left(V_{\max}\delta\right)^2, n = 1, ..., N-1. \tag{14h}$$

where (14b) and (14c) denote the QoS constraint. Constraints (14d) and (14e) shows that each subcarrier can be allocated to only one user at each time slot. (14f)- (14h) implies the UAV fly trajectory constraint. It can be seen that the problem P0 is a nonconvex optimization problem due to the highly coupled optimization variables. In addition, it is a mixed integer non-convex optimization problem arises from containing binary variable. Therefore, to solve the above difficulties, we propose an efficient iterative algorithm in the next section to attain a suboptimal solution.

3. PROBLEM SOLUTION

In this section, we divide problem P0 into two sub-problems, and propose a computationally efficient iterative alternating algorithm to achieve a suboptimal solution. In particular, we first optimize resource allocation for a given UAV trajectory and then optimize the UAV's trajectory for a given resource allocation scheme. The two optimization problems are solved alternately until convergence is achieved.

3.1. Optimization Resource of Allocation

In this section, we consider sub-problem P1 for optimizing the resource allocation B for given UAV trajectory Q, which can be written as

$$\text{P1} \qquad \max_{\boldsymbol{B}} \tilde{S}_{sum}, \tag{15a}$$

$$s.t. \qquad \sum_{i=1}^{I_F} b_k^i[n] \frac{I}{OL} \varepsilon_O \ge S_{th}^k, \forall n, k, \tag{15b}$$

$$\sum_{k=1}^{K} b_k^i[n] \le 1, \forall n, i, \tag{15c}$$

$$b_k^i[n] \in \{0,1\}, \forall n, k, i, \tag{15d}$$

Since the challenge of solving the optimization problem P1 lies on the 0-1 constraint in (15d), we first relax b into the continuous variable between 0 and 1. Then, the obtained solution from the constraint relaxed problem will serve as a building block for the development of the optimal solution of the original problem. Thus, the sub-problem P1 is reformulated as

$$\text{P1.1} \qquad \max_{\boldsymbol{B}} \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{i=1}^{I_F} b_k^i[n] \frac{I}{OL} \tilde{\varepsilon}_O \tag{16a}$$

$$s.t. \qquad 0 < b_k^i[n] \le 1, \forall n, k, \tag{16b}$$

$$\text{(15a), (15b)}$$

Furthermore, it can be verified that P1.1 satisfies the Slater's constraint qualification. Therefore, the strong duality holds and the duality gap is zero, which means the dual problem is equivalent to solving the primal problem of P1.1. Hence, we use the Lagrange dual method [32] to obtain its dual optimization problem herein. By introducing the non-negative Lagrange multiplier $\{\alpha\} \triangleq \{\alpha_{n,k}, \forall n,k\} \geq 0$, and $\{\beta\} \triangleq \{\beta_{n,i}, \forall n,i\} \geq 0$ corresponding to constraints (15b) and (15c) respectively, we derive the Lagrange function of P1.1.

$$L(\alpha,\beta,b) = \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{i=1}^{I_F}\left(\left(\alpha_{n,k}+1\right)S_k[n]-\beta_{n,i}\right)b_k^i[n] - \sum_{n=1}^{N}\sum_{k=1}^{K}\alpha_{n,k}S_{th}^k + \sum_{n=1}^{N}\sum_{i=1}^{I_F}\beta_{n,i}, \forall n,k,i, \tag{17}$$

The boundary constraints will be absorbed into the optimal solution in the following. Then, the Lagrange dual function which is defined as the maximum value (17) over b is given by

$$f(\alpha,\beta) = \max_{\boldsymbol{b}} L(\alpha,\beta,b) \tag{18}$$

$$s.t. \qquad 0 < b_k^i[n] \leq 1, \forall n,k, \tag{18a}$$

for which the following lemma 2 holds.

Lemma 1 To make the Lagrange dual function $f(\alpha,\beta)$ is bounded, i.e, $f(\alpha,\beta) < \infty$, it must follows

$$\left(\alpha_{n,k}+1\right)S_k[n]-\beta_{n,i} = 0, \forall n,k,i. \tag{19}$$

Proof If $\left(\alpha_{n,k}+1\right)S_k[n]-\beta_{n,i} < 0$ (or $\left(\alpha_{n,k}+1\right)S_k[n]-\beta_{n,i} > 0$), it follows that the Lagrange dual function $f(\alpha,\beta) \to \infty$ by letting $b \to +\infty$ (or $b \to -\infty$), which is contradictory with the boundedness of $f(\alpha,\beta)$. Thus the lemma is proved.

Hence, the dual problem of problem P1.1 can be written as

$$\min_{\alpha,\beta\geq 0} \max_{\boldsymbol{b}} L(\alpha,\beta,b) \tag{20}$$

$$s.t. \qquad (19)$$

Now, the constraints in the dual problem only holds when $\beta_{n,i} = \beta_{n,j} > 0$, $\forall n,i,j$. It means that all the subcarriers share the same value of the Lagrange multiplier for constraint (15c). In fact, the channel from the UAV to user node k are LoS-dominated, which leads to a frequency-flat fading across all the subcarriers. The second important observation is that, with $S_k[n] \neq S_{k'}[n]$, $\forall n,k,k'$, at most one α_{n,k^*} can be zero with $k^* = \max_k R_k[n]$ at each time slot n.

In particular, if $\alpha_{n,k} = \alpha_{n,k'} = 0, \forall n$, we have $S_k[n] = S_{k'}[n] = \beta_{n,i}$ based on (19), which leads to a contradiction. On the other hand, if $S_k[n] > S_{k'}[n]$ with $\alpha_{n,k} > 0$ and $\alpha_{n,k'} = 0$, we have $\left(\alpha_{n,k}+1\right)S_k[n] = S_{k'}[n] = \beta_{n,i}$, which also leads to a contradiction. Therefore, the Lagrange multiplier of the minimum rate constraint (15b) for the strongest user $k^* = \max_k S_k[n]$ is the only one to be zero at the optimal point.

The physical insight of this observation can be revealed based on the Karush-Kuhn-Tucker (KKT) conditions. According to the KKT conditions, the following equation should hold at the optimal point of the problem in (16).

$$\alpha_{n,k}\left(\sum_{i=1}^{I_{\mathrm{F}}} b_k^i\left[n\right] S_k\left[n\right]-S_{th}^k\right)=0,\forall n,k. \tag{21}$$

Therefore, if $\alpha_{n,k}>0$, the minimum data rate requirement (15b) of user k at time slot n must be satisfied with equality. On the other hand, if $\alpha_{n,k'}=0$, i.e., user k is the strongest user at time slot n, we have $\sum_{i=1}^{I_{\mathrm{F}}} b_k^i\left[n\right] S_k\left[n\right]\geq S_{th}^k$. The physical meaning of this observation is that no exceedingly large amount of resources should be allocated to any users except for the strongest user k^* when user k, $\forall k\in\{\mathrm{K}/k^*\}$, such that user k satisfies constraint (15b) with equality.

Table 1 Optimal Resource Allocation for Given the Trajectory of UAV

Algorithm 1: Optimal Resource Allocation for Given the Trajectory of UAV

Input: For a user set $k\in\mathrm{K}\triangleq\{1,\cdots,K\}$ and a given UAV trajectory;

for $n=1,2,\mathrm{K},N$ **do**

Find the strongest user $k^*=\max\limits_{\mathrm{K}} S_k[n]$ at time slot n;

Compute the minimum number of subcarriers for all users according to $I_k[n]=\left\lceil S_{th}^k/S_k[n]\right\rceil,\ \forall k\in\{\mathrm{K}/k^*\}$;

Compute the number of subcarriers for the strongest user by: $I_{k^*}[n]=I_F-\sum_{k\in\{\mathrm{K}/k^*\}} I_k[n]$;

if $I_{k^*}[n]S_{k^*}[n]\geq S_{th}^k$ **then**

Set $i=1$;

Allocate $I_k[n]$ subcarriers to user k.

for $k=1:K$ **do**

for $i_{\text{index}}=1:I_k[n]$ **do**

$s_k^i[n]=1$;

$i=i+1$;

end for

end for

end if

Stop and declare problem P1.1 infeasible

end for

Now, due to the binary constraint (15c), the minimum data rate requirement (15b) of user k at time slot n may not hold with equality at the optimal point. However, based on the insight from (20) and (21), we can propose a sequentially subcarrier allocation algorithm summarized in algorithm 1. Note that $\lceil\cdot\rceil$ is the ceiling function which returns the smallest integer greater than the input value.

3.2. Optimization of UAV Trajectory

In this section, we consider sub-problem 2 for optimizing UAV trajectory by assuming that the resource allocation is fixed, which yields

$$\mathrm{P2}\qquad \max_{\boldsymbol{Q}}\ \tilde{S}_{sum} \tag{22a}$$

$$s.t.\qquad \tilde{S}_k[n]\geq S_{th}^k,\forall n,k, \tag{22b}$$

$$\tilde{\varepsilon}_O\left(\gamma\right)\geq\varepsilon_{th}^k,\forall k\ , \tag{22c}$$

$$\mathbf{q}_I = \mathbf{q}[1], \tag{22f}$$

$$\mathbf{q}_F = \mathbf{q}[N], \tag{22g}$$

$$\| \mathbf{q}[n+1] - \mathbf{q}[n] \|^2 \le (V_{\max}\delta)^2, n = 1, ..., N-1. \tag{22h}$$

By introducing slack variables $V = \{v_k[n], \forall k, n\}$, the problem P2 is converted into

$$\text{P2.1} \qquad \max_{\boldsymbol{Q},\boldsymbol{V}} \sum_{n=1}^{N} \sum_{k=1}^{K} b_k^i[n] \frac{I}{OL} (A_{O,1} + \frac{A_{O,2} - A_{O,1}}{1 + e^{-(C_{O,1} v_k[n] + C_{O,2})}}) \tag{23a}$$

$$s.t. \qquad \sum_{i=1}^{I_F} b_k^i[n] \frac{I}{OL} \tilde{\varepsilon}_O(v_k[n]) \ge S_{th}^k, \forall n, k, \tag{23b}$$

$$v_k[n] \le \frac{\gamma_0 P}{\| \mathbf{q}[n] - \mathbf{w}_k \|^2 + H^2}, \forall k, n, \tag{23c}$$

$$(22\text{f}) - (22\text{h})$$

First, we can observe that the constraint (23b) and (23c) is neither concave nor convex w.r.t. the optimization variables $v_k[n]$ and $\mathbf{q}[n]$. To tackle this difficulty, we first introduce an important lemma as below. we can easily prove that $\tilde{S}_k[n]$ and $v_k[n]$ is a convex function w.r.t. $(1 + e^{-(C_{O,1} v_k[n] + C_{O,2})})$ and $(\| \mathbf{q}[n] - \mathbf{w}_k \|^2 + H^2)$. Therefore, we can leverage the SCA technique to derive its convex approximation. To be specific, using the fact that the first-order Taylor approximation of a convex function is a global under-estimator, we obtain the lower bound of $\tilde{S}_k[n]$ and $v_k[n]$ in the r-th iteration with given local point $\hat{\mathbf{q}}[n]$.

$$\tilde{S}_k[n] \ge \tilde{S}_k^{\text{lb}}[n] \triangleq \hat{S}_k[n] - \hat{\Phi}_k[n](e^{-l_k[n]} - e^{-\hat{l}_k[n]}), \forall k, n, \tag{24}$$

$$v_k[n] \ge v_k^{\text{lb}}[n] \triangleq \hat{v}_k[n] - \hat{\Lambda}_k[n](\| \boldsymbol{q}[n] - \boldsymbol{w}_k \|^2 - \| \hat{\boldsymbol{q}}[n] - \boldsymbol{w}_k \|^2), \forall k, n, \tag{25}$$

where

$$\hat{S}_k[n] = (A_{O,1} + \frac{A_{O,2}}{1 + e^{-C_{O,1}\hat{v}_k[n] + C_{O,2}}}), \tag{26}$$

$$\hat{\Phi}_k[n] = \frac{A_{O,2}}{\left(1 + e^{-C_{O,1}\hat{v}_k[n] + C_{O,2}}\right)^2}, \tag{27}$$

$$l_k[n] \triangleq C_{O,1} v_k[n] + C_{O,2}, \tag{28}$$

$$\hat{v}_k[n] = \frac{\gamma_0}{\| \hat{\boldsymbol{q}}[n] - \boldsymbol{w}_k \|^2 + H^2}, \tag{29}$$

$$\hat{\Lambda}_k[n] = \frac{\gamma_0}{(\| \hat{\boldsymbol{q}}[n] - \boldsymbol{w}_k \|^2 + H^2)^2}. \tag{30}$$

Consequently, problem P2.1 can be transformed to the following approximate problem by substituting (24) and (25) into (23b) and (23c), respectively.

$$\text{P2.2} \qquad \max_{\boldsymbol{q_n}, \mathrm{l}_{\boldsymbol{k}}[\boldsymbol{n}]} \sum_{n=1}^{N} \sum_{k=1}^{K} b_k^i[n] \frac{I}{OL} \tilde{S}_k^{\text{lb}}[n] \tag{31a}$$

$$s.t. \quad \sum_{i=1}^{I_F} b_k^i[n] \frac{I}{OL} \tilde{\varepsilon}_k^{\mathrm{lb}}[n] \geq S_{th}^k, \forall n, k, \tag{31b}$$

$$l_k[n] \leq B_1 + B_2 v_k^{\mathrm{lb}}[n], \ \forall k, n, \tag{31c}$$

$$(22\mathrm{f}) - (22\mathrm{h})$$

Problem P2.2 is now a convex optimization problem, which can be efficiently solved by using existing solvers, e.g., CVX. It is worthwhile to note that, by approximating the convex constraints with their convex lower bounds, the feasible set of problem P2.2 is always a subset of problem P2.1. Therefore, solving problem P2.2 gives the lower bound of the objective value in problem P2.1.

3.3. Joint Optimization Algorithm Design

Based on the above solution and theoretical analysis, an efficient iterative alternating algorithm is proposed in this paper, and the algorithm flow is summarized in Algorithm Table 3-2.

Table 2 Iterative Alternating Resource Allocation and Trajectory Optimization

Algorithm 2: Iterative Alternating Resource Allocation and Trajectory Optimization	
1:	Initialize the maximum number of iterations L, iteration index $l' = 0$, and resource allocation policy as s(0)
2:	**repeat**
3:	For the fixed resource allocation $s^{(l')}$, obtain the optiminal trajctory (x, y);
4:	For the fixed UAV's trajectory (x, y), obtain the intermediate optimal resource allocation using Algorithm 1.
5:	set $l' = l' + 1$ and $s^{(l')} = s$, $(x, y) = \left(x^{(l')}, y^{(l')}\right)$
6:	**until** convergence or iteration index reaches to the maximum number
7:	$s^* = s^{(l')}$, $\left(x^*, y^*\right) = \left(x^{(l')}, y^{(l')}\right)$.

3.4. Algorithm convergence

In summary, the proposed algorithm solves the two sub-problems P1 and P 2 in an alternating manner. Since the objective value of P0 with the solutions obtained by solving sub-problems P1 and P2 is non-decreasing over iteration, and the optimal value of P0 is finite, the solution obtained by the proposed algorithm is guaranteed to converge to a suboptimal solution. Furthermore, it can be shown that the proposed algorithm is guaranteed to converge to a solution that satisfies at least the KKT condition of the original optimization problem P0 by following a similar argument in the literature [33].

4. SUMMARY

This paper investigated a UAV-enabled downlink semantic communication system providing communication services to ground users. A joint trajectory and resource allocation algorithm was proposed to maximize the semantic information sum rate subject to UAV's mobility constraints, the minimum required semantic data rate for each user node and the minimum semantic similarity that can be received. And we also further proofs demonstrate that the proposed algorithm is guaranteed to converge to a solution that satisfies at least the KKT condition of the original optimization problem.

ACKNOWLEDGEMENTS

This work is partially supported by the National Natural Science Foundation of China (NSFC) under Grants No. 61801171 and 62072158, in part by the Henan Province Science Fund for Distinguished Young Scholars (222300420006), and Program for Innovative Research Team in University of Henan Province (21IRTSTHN015).

REFERENCES

[1] Liang RJ, Fan JC. Energy-Efficient mmWave IoT Communications with multi-hop IRS-Assisted Systems[J]. IEEE Internet of Things Journal, 2023, Vol. 10 (No. 21), p. 19344-19355.

[2] M. S. Abouamer, P. Mitran. Joint uplink-downlink resource allocation for multiuser irs-assisted systems[J]. IEEE Transactions on Wireless Communications, 2023, Vol. 21 (No. 12), p. 10918-10933.

[3] M. Matin, S. K. Goudos, S. Wan, et al. Artificial intelligence (AI) and machine learning (ML) for beyond 5g/6g communications[J]. EURASIP Journal on Wireless Communications and Networking, 2023, Vol. 1, p. 22.

[4] W. Saad, M. Bennis, M. Chen. A vision of 6g wireless systems: Applications, trends, technologies, and open research problems[J]. IEEE Network, 2020, Vol. 34 (No. 3), p. 134-142.

[5] C. Munasinghe, F. M. Amin, D. Scaramuzza, et al. Covered, collaborative robot environment dataset for 3d semantic segmentation[J]. arXiv preprint arXiv:2302.12656, 2022.

[6] A. Cavagna, N. Li, A. Iosifidis, et al. Semantic communication enabling robust edge intelligence for time critical iot applications[J]. arXiv preprint arXiv:2211.13787, 2022.

[7] J. Yu, A. Alhilal, P. Hui, et al. 6g mobile-edge empowered metaverse: Requirements, technologies, challenges and research directions[J]. arXiv preprint arXiv:2211.04854, 2022.

[8] C. E. Shannon. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, Vol. 27 (No. 03), p. 379-423.

[9] W. Yang, H. Du, Z. Q. Liew, et al. Semantic communications for future internet: Fundamentals, applications, and challenges[J]. IEEE Communication Survey Tutorials, 2023, Vol. 25 (No. 01), p. 213-250.

[10] H. Xie and Z. Qin. A lite distributed semantic communication system for internet of things[J]. IEEE Journal of Selected Areas in Communications, 2021, Vol. 39 (No. 01), p. 142-153, 2021.

[11] K. Lu, R. Li, X. Chen, et al. Reinforcement learning-powered semantic communication via semantic similarity, CoRR, vol. abs/2108.12121, 2021.

[12] Q. Zhou, R. Li, Z. Zhao, et al. Semantic communication with adaptive universal transformer[J]. IEEE Wireless Communication Letters, 2022, Vol. 11 (No. 3), p. 453-457.

[13] E. C. Strinati and S. Barbarossa. 6g networks: Beyond shannon towards semantic and goal-oriented communications[J]. Computer Networks, 2021, Vol. 190, p. 107930.

[14] L. Xia, Y. Sun, X. Li, et al. Wireless resource management in intelligent semantic communication networks[C]. IEEE Conference on Computer Communications Workshops, INFOCOM 2022 Workshops, New York, NY, USA, May 2-5, 2022, p. 1-6.

[15] N. Farsad, M. Rao, A. Goldsmith. Deep learning for joint source-channel coding of text[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, p. 2326-2330.

[16] H. Xie, Z. Qin, G. Y. Li, et al. Deep learning enabled semantic communication systems[J]. IEEE Transactions Signal Process, 2021, Vol. 69, p. 2663–2675.

[17] X. Kang, B. Song, J. Guo, et al. Task oriented image transmission for scene classification in unmanned aerial systems[J]. IEEE Transactions Communication, Vol. 70 (No. 8), p. 5181-5192.

[18] L. Yan, Z. Qin, R. Zhang, et al. Resource allocation for text semantic communications[J]. IEEE Wireless Communication. Letters, 2022, Vol. 11 (No. 07), p. 1394-1398.

[19] K. Meng, Q. Wu, S. Ma, et al. UAV trajectory and beamforming optimization for integrated periodic sensing and communication[J]. IEEE Wireless Communication Letters, 2022, Vol. 11 (No. 06), p. 1211-1215.

[20] S. Li, N. Zhang, H. Chen, et al. Joint subcarrier allocation, modulation mode selection, and trajectory design in a UAV-based OFDMA network[J]. IEEE Communication Letters, 2022, Vol. 26 (No. 09), p. 111-2115.

[21] Y. Liu, B. Duo, Q. Wu, et al. Full-dimensional rate enhancement for uav-enabled communications via intelligent omni-surface[J] IEEE Communication Letters, 2022, Vol. 11 (No. 09), p. 1955-1959.

[22] Y. Zeng, R. Zhang, and T. J. Lim. Throughput maximization for uav-enabled mobile relaying systems[J]. IEEE Transactions Communications, 2016, Vol. 64 (No. 12), p. 4983-4996.

[23] S. Zeng, H. Zhang, B. Di, et al. Trajectory optimization and resource allocation for OFDMA UAV relay networks[J]. IEEE Transactions Wireless Communications, 2021, Vol. 20 (No. 10), p. 6634–6647.

[24] J. Kang, H. Du, Z. Li, et al. Personalized saliency in task-oriented semantic communications: Image transmission and performance analysis[J]. IEEE Journal of Selected Areas in Communications, Vol. 41 (No. 1), p. 186-201.

[25] L. Xia, Y. Sun, D. Niyato, et al. Wireless semantic communication: A networking perspective[J]. arXiv preprint arXiv:2212.14142.

[26] C. Liu, C. Guo, Y. Yang, et al. Bandwidth and power allocation for task-oriented semantic communication[J]. arXiv preprint arXiv:2201.10795 2022.

[27] Y. Wang, M. Chen, T. Luo, et al. Performance optimization for semantic communications: An attention-based reinforcement learning approach[J]. IEEE Journal of Selected Areas in Communications, 2022, Vol. 40 (No. 9), p. 2598-2613.

[28] C. Liu, C. Guo, Y. Yang, et al. Adaptable semantic compression and resource allocation for task-oriented communications[J]. arXiv preprint arXiv: 2204.08910.

[29] Z. Q. Liew, Y. Cheng, W. Y. B. Lim, et al. Economics of semantic communication system in wireless powered internet of things[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2022, Virtual and Singapore, 23-27 May 2022, p. 8637-8641.

[30] L. Yan, Z. Qin, R. Zhang, et al. QoE-based semantic-aware resource allocation for multi-task networks[J]. arXiv preprint arXiv:2305.06543.

[31] X. Mu, Y. Liu, L. Guo, et al. Heterogeneous semantic and bit communications: A semi-noma scheme[J]. IEEE Journal of Selected Areas in Communications, 2023, Vol. 40 (No. 1), p. 155-169.

[32] S. H. Low and D. E. Lapsley. Optimization flow controli: basic algorithm and convergence[J]. IEEE/ACM Transactions Networks, 1999, Vol. 7 (No. 6), p. 861-874.