Performance Optimization for Wireless Semantic Communications over Energy Harvesting Networks

Mingzhe Chen, Yining Wang, and H. Vincent Poor,

***Abstract*—In this paper, the optimization of semantic communications over energy harvesting networks is studied. In the considered model, a set of users use semantic communication techniques and the harvested energy to transmit text data to a base station (BS). Here, semantic communication techniques enable each user to transmit the meaning of the original data (called semantic information) thereby reducing its transmission delay and energy consumption. The BS will recover the data using the received semantic information. To further improve communication efficiency, each user can transmit only partial semantic information to the BS. Therefore, each user needs to jointly determine the partial semantic information to be transmitted and the resource block (RB) that is used for semantic information transmission. This problem is formulated as an optimization problem whose goal is to maximize the sum of all users' similarities that capture the differences between the original data that each user needs to transmit and the data recovered by the BS. To solve this problem, a value decomposition based deep Q network is proposed, which enables the users to jointly find the semantic information transmission and the RB allocation schemes that maximize the sum of all users' similarities. Simulation results demonstrate that the proposed method can improve sum of all users' similarities by up to three-fold compared to the independent reinforcement learning.**

I. Introduction

Due to advances in microelectronics, contemporary wireless edge devices have computational power to generate, process, and transmit a large amount of data. Therefore, one can use semantic communication techniques [1] that enable wireless edge devices to extract and transmit the meaning of original data (called semantic information) to reduce the heavy congestion of current wireless networks thus improving communication efficiency. However, using semantic communication techniques for data transmission faces several challenges such as semantic information modeling and extraction, original data recovery, and the design of semantic metrics that can capture the effects of wireless factors (e.g., transmit power, packet errors) on the semantic communications.

Several existing works [1]–[5] have studied problems related to semantic communications over wireless networks. In particular, the authors in [1] and [2] have provided comprehensive reviews on the existing semantic communication systems and pointed out the challenges remaining for the use of semantic communication techniques for wireless edge devices. Further, the work in [3] provides a review of the application of federated learning to semantic communications. However, the works in [1]–[3] do not provide rigorous mathematical models and analysis. In [4], the authors designed a deep learning based semantic communication system for speech signals and used an attention mechanism to improve the training of machine learning models. The work in [5] investigated a semantic communication scenario where a set of users transmit correlated data to a base station (BS) and proposed a deep neural network enabled multi-user semantic communication system that can extract the semantic information of image and text from different users. However, the authors in [4] and [5] did not extract the practical meaning from the original data since they used a neural network as an encoder to extract the meaning of the original data and considered the output of the encoder to be the data's meaning. Further, the authors in [1]–[5] did not jointly optimize the wireless resources (e.g., transmit power and bandwidth) and semantic information transmission. In particular, to achieve transmission constraints (e.g., delay and data rate requirements), one can transmit only partial semantic information to other devices. In addition, none of the works [1]–[5] considered the optimization of semantic communication over energy harvesting networks. In an energy harvesting network, each user can use the harvested energy for semantic information transmission. However, each user may not know the energy that it can harvest in advance and hence it may not be able to predetermine the amount of semantic information to be transmitted.

The main contribution of this work is a novel framework that enables a set of users to communicate with a BS using semantic communication and energy harvesting techniques. In particular, we consider an energy harvesting network in which a set of users use semantic communication techniques and the energy that is harvested from their own sources to transmit text data to the BS. The BS can recover the data using the received semantic information. Due to limited spectrum resources and energy, each user may not be able to transmit all of its semantic information to the BS. Therefore, each user needs to determine the amount of semantic information to be transmitted and the resource block (RB) used to transmit the semantic information. We formulate this problem as an optimization problem that aims to maximize the sum of all users' similarities between the original data and the recovered data. To solve this problem, we use a value decomposition based deep Q network (VD based DQN) that enables the users to maximize the sum of all users' similarities via individually determining the amount of semantic information to be transmitted and the RB used for this information transmission. Meanwhile, the training of VD based DQN requires each user to share only its reward and Q function values thus reducing the complexity and communication overhead of training the VD based DQN.

This work was supported in part by the National Science Foundation under Grants CCF-1908308 and CNS-2128448.

M. Chen and H. V. Poor are with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ, 08544, USA (Emails: mingzhec@princeton.edu; poor@princeton.edu).

Y. Wang is with the Beijing Laboratory of Advanced Information Network, Beijing University of Posts and Telecommunications, Beijing, 100876, China (Email: wyy0206@bupt.edu.cn).

II. System Model and Problem Formulation

Consider a cellular network in which a set $\mathcal{U}$ of U users that are powered by its own energy harvesting source transmit their text data to a BS using semantic communication techniques. Here, semantic communication techniques enable the users to transmit the meaning of their data to the BS which will recover the original data according to its received data. Hereinafter, the meaning that each user extracts from its original data is called *semantic information*. Next, we introduce the models for semantic information extraction, semantic information transmission, and original data recovery.

A. Semantic Information Extraction

Let $k_{i,n}$ be a token n of the text data of user i, which represents a word, a symbol, or a punctuation. The data that user i needs to transmit to the BS is $\boldsymbol{k}_i = \{k_{i,1}, \ldots, k_{i,N_i}\}$ where N_i is the number of tokens in the original data of user i. To model the semantic information extracted from $\boldsymbol{k}_i$, we use a knowledge graph (KG) [6]. Therefore, the semantic information of $\boldsymbol{k}_i$ is

$$\mathcal{S}_i = \left\{ \boldsymbol{s}_i^1, \ldots, \boldsymbol{s}_i^j, \ldots, \boldsymbol{s}_i^{G_i} \right\}, \tag{1}$$

where $\boldsymbol{s}_i^j = [\boldsymbol{e}_{i,m}^j, \boldsymbol{r}_{i,mn}^j, \boldsymbol{s}_{i,n}^j]$ with $\boldsymbol{e}_{i,m}^j$ being a node m, that consists of $Z\left(\boldsymbol{e}_{i,m}^j\right)$ tokens in $\boldsymbol{k}_i$; $\boldsymbol{e}_{i,m}^j$ represents a concept or an object in the real word; $\boldsymbol{r}_{i,mn}^j$ is the relationship between node m and node n, which consists of $Z\left(\boldsymbol{r}_{i,mn}^j\right)$ tokens in $\boldsymbol{k}_i$. Here, $\boldsymbol{r}_{i,mn}^j$ is directional and hence, we have $\boldsymbol{r}_{i,mn}^j \neq \boldsymbol{r}_{i,nm}^j$; G_i is the number of elements in $\mathcal{S}_i$. For example, let $\boldsymbol{e}_{i,m}^j = [\text{stochastic}, \text{model}]$, $\boldsymbol{e}_{i,n}^j = [\text{speech}, \text{identification}]$, and $\boldsymbol{r}_{i,mn}^j = [\text{used}, \text{for}]$. We have $Z\left(\boldsymbol{e}_{i,m}^j\right) = Z\left(\boldsymbol{r}_{i,mn}^j\right) = 2$. Therefore, the total number of tokens used to represent the semantic information $\mathcal{S}_i$ is

$$Z(\mathcal{S}_i) = \sum_{j=1}^{G_i} \left(Z\left(\boldsymbol{e}_{i,m}^j\right) + Z\left(\boldsymbol{e}_{i,n}^j\right) + Z\left(\boldsymbol{r}_{i,mn}^j\right) \right). \tag{2}$$

From (2), we see that the number of tokens, $Z(\mathcal{S}_i)$, used to represent the semantic information is smaller than the number of tokens, N_i, used to represent the original data. Therefore, for each user i, transmitting semantic information can reduce the size of data needed to transmit over wireless links.

B. Semantic Information Transmission

Next, we introduce the model of semantic information transmission of each user i.

1) Semantic Information Transmission Model: An orthogonal frequency division multiple access (OFDMA) technique is used for semantic information transmission. We assume that a set $\mathcal{R}$ of R RBs can be allocated to the users and each user can only occupy one RB [7]. Hence, the data rate of each user i transmitting the semantic information to the BS is

$$c_i(\boldsymbol{r}_{i,t}) = \sum_{q=1}^{R} r_{i,t}^q B \log_2 \left(1 + \frac{P h_i^q}{I_q + B N_0} \right), \tag{3}$$

where $\boldsymbol{r}_{i,t} = \left[r_{i,t}^1, \ldots, r_{i,t}^R \right]$ with $r_{i,t}^q \in \{0, 1\}$; $r_{i,t}^q = 1$ implies that RB q is allocated to user i at time slot t, otherwise, we have $r_{i,t}^q = 0$; h_i is channel gain which depends on the location of user i at time slot t; B is the bandwidth of each RB; I_q is the interference over RB q. Here, I_q is the interference caused by other users that occupy the same RB but associated with other BSs; N_0 is the noise power spectral density.

Due to limited bandwidth and energy, each device may not be able to transmit all the semantic information to the BS. Therefore, we assume that each user can transmit a certain amount of the semantic information to the BS. The amount of semantic information that each user i needs to transmit at time slot t is $\mathcal{S}_{i,t}'$ with $\mathcal{S}_{i,t}' \subseteq \mathcal{S}_{i,t}$. Then, the time that user i uses to transmit $\mathcal{S}_{i,t}'$ is

$$\psi\left(\boldsymbol{r}_{i,t}, \mathcal{S}_{i,t}'\right) = \frac{Z\left(\mathcal{S}_{i,t}'\right) O}{c_i\left(\boldsymbol{r}_{i,t}\right)}, \tag{4}$$

where O is the number of bits used to represent each token. The energy that each user i uses to transmit the semantic information is $P\psi\left(\boldsymbol{r}_{i,t}, \mathcal{S}_{i,t}'\right)$.

C. Original Data Recovery

The BS will use the received semantic information to recover the original text data. Here, we use a well trained graph-to-text generation model [8] to recover the original data and the recovered text is $\boldsymbol{k}_i'\left(\mathcal{S}_{i,t}'\right)$. To measure the quality of the recovered text data, we first use a deep neural network (DNN) to vectorize the original text data $\boldsymbol{k}_{i,t}$ and the recovered data $\boldsymbol{k}_i'\left(\mathcal{S}_{i,t}'\right)$. In particular, we assume that the vectorized original text data is $\boldsymbol{k}_i^{\mathrm{V}} = \{k_{i,1}^{\mathrm{V}}, \ldots, k_{i,N_i^{\mathrm{V}}}^{\mathrm{V}}\}$ and the vectorized recovered data is $\boldsymbol{k}_i^{\mathrm{V}'}\left(\mathcal{S}_{i,t}'\right)$.

A cosine similarity function [9] is used to capture the differences between the meaning of the vectorized original text data $\boldsymbol{k}_i^{\mathrm{V}}$ and the meaning of the vectorized recovered data $\boldsymbol{k}_i^{\mathrm{V}'}\left(\mathcal{S}_{i,t}'\right)$, as follows:

$$M\left(\mathcal{S}_{i,t}'\right) = \frac{\boldsymbol{k}_i^{\mathrm{V}} \cdot \boldsymbol{k}_i^{\mathrm{V}'}\left(\mathcal{S}_{i,t}'\right)^{\mathrm{T}}}{\left\|\boldsymbol{k}_i^{\mathrm{V}}\right\| \times \left\|\boldsymbol{k}_i^{\mathrm{V}'}\left(\mathcal{S}_{i,t}'\right)\right\|}. \tag{5}$$

D. Problem Formulation

Next, we introduce our optimization problem whose goal is to maximize the similarities of all users while satisfying the transmission delay and energy consumption requirements. The studied optimization problem includes determining the amount of semantic information that needs to transmit and optimizing RB allocation for each user with the uncertain mobility pattern and harvested energy. We assume that the initial energy of each user i at time slot 0 is E_i^{O}. We also assume that each user can harvest E_i^t energy at time slot t where E_i^t follows an unknown distribution. The similarity optimization problem over a set $\mathcal{T}$ of T time slots is

$$\max_{\left\{\boldsymbol{r}_{i,t}, \mathcal{S}_{i,t}'\right\}_{\substack{i \in \mathcal{U} \\ t \in \mathcal{T}}}} \sum_{t=1}^{T} \sum_{i=1}^{U} M\left(\mathcal{S}_{i,t}'\right), \tag{6}$$

$$\text{s. t. } r_{i,t}^{q}, \in \{0,1\}, \quad \forall i \in \mathcal{U}, t \in \mathcal{T}, \tag{6a}$$

$$\sum_{i \in \mathcal{U}} r_{i,t}^{q} \leq 1, \ \forall q \in \mathcal{R}, t \in \mathcal{T}, \tag{6b}$$

$$\sum_{q \in \mathcal{R}} r_{i,t}^{q} \leq 1, \ \ \forall i \in \mathcal{N}, t \in \mathcal{T}, \tag{6c}$$

$$E_i^{\mathrm{O}} + \sum_{j=1}^{t} E_i^j - P \sum_{j=1}^{t} \left(\sum_{q \in \mathcal{R}} r_{i,j}^{q} \right) \times \psi\left(\boldsymbol{r}_{i,t}, \boldsymbol{\varphi}_{i,t}, \mathcal{S}_{i,t}' \right) \geq 0, \quad \forall i \in \mathcal{N}, \ t \in \mathcal{T}, \tag{6d}$$

where $\sum_{j=1}^{t} E_i^j$ is the sum of the harvested energy of user i from time slot 1 to time slot t, $P \sum_{j=1}^{t} \left(\sum_{q \in \mathcal{R}} r_{i,j}^{q} \right) \times \psi\left(\boldsymbol{r}_{i,t}, \boldsymbol{\varphi}_{i,t}, \mathcal{S}_{i,t}' \right)$ is the sum of energy consumption of user i from time slot 1 to time slot t where $\sum_{q \in \mathcal{R}} r_{i,j}^{q}$ implies whether user i transmit the semantic information to the BS. Constraints (6a) and (6b) implies that each RB can be allocated to only one user at each time slot. Constraint (6c) indicates that each user can only occupy one RB at each time slot. Constraint (6d) implies that each user must have enough energy for semantic information transmission. The problem (6) is challenging to solve due to the following reasons. First, since the semantic information extraction and original text recovery are performed by deep neural networks, we cannot find a close-form expression to show the relationship between original text data $\boldsymbol{k}_{i,t}$ and recovered text data $\boldsymbol{k}_i'\left(\mathcal{S}_{i,t}'\right)$. Therefore, we may not be able to use the methods (e.g., optimizaiton theory) that need to know how the optimization variables affect the objective function to solve problem (6). Meanwhile, the BS does not know either the harvested energy of each user or the text data that each user needs to transmit. Hence, it is impractical to use a centralized algorithm to solve problem (6). In consequence, we propose the use of a distributed reinforcement learning (DRL) algorithm to solve problem (6). Compared to centralized learning algorithms, DRL enables each users to individually determine its semantic information that will be transmitted and the BS determine the optimal RB allocation scheme for the users. Hence, using DRL, the users and the BS can jointly maximize the similarities of all users while meeting the energy constraint with less communication overhead.

III. Value Decomposition Based DRL

Value decomposition based DRL [10] is a cooperative DQN algorithm which enables the users and the BS to jointly find team RB allocation and partial semantic information transmission schemes so as to optimize the similarities of all users. In other words, VD based DQN can find the RB allocation and partial semantic information transmission schemes to maximize the sum of all user's similarities. However, each individual user's similarity may not be maximized. Next, we first introduce the components of VD based DQN. Then, we introduce the training process of VD based DQN.

A. Components of VD based DQN

VD based DRL is a policy gradient based RL algorithm. Hence, it consists of five components:

- *Agents*: The agents that perform VD based DRL is the users. The BS will calculate the parameters that will be used for training VD-based DQN. However, it will not implement VD based DQN.
- *States*: A state of each agent is used to describe the unique environment of each user. In particular, a state of each user consists of the location and the remaining energy of each user. Hence, a state of user i at time slot t is $\boldsymbol{\xi}_{i,t} = [\boldsymbol{\varphi}_i, b_{i,t}]$, with $\boldsymbol{\varphi}_i$ being the location of user i and $b_{i,t} = E_i^{\mathrm{O}} + \sum_{j=1}^{t-1} E_i^j - P \sum_{j=1}^{t-1} \left(\sum_{q \in \mathcal{R}} r_{i,j}^{q} \right) \times \psi\left(\boldsymbol{r}_{i,t}, \boldsymbol{\varphi}_{i,t}, \mathcal{S}_{i,t}' \right)$. The states of all agents at time slot t can be represented by a vector $\boldsymbol{\xi}_t = \left[\boldsymbol{\xi}_{1,t}, \ldots, \boldsymbol{\xi}_{U,t}\right]$.
- *Actions*: The action of each agent is the energy and the RB that each user i use for partial semantic information transmission. Hence, an action of user i at time slot t can be expressed as $\boldsymbol{a}_{i,t} = [\eta_{i,t}, \boldsymbol{r}_{i,t}]$, where $\eta_{i,t}$ is the energy that user i uses to transmit the semantic information at time slot t. Similarly, the actions of all users at time t is $\boldsymbol{a}_t = [\boldsymbol{a}_{1,t}, \ldots, \boldsymbol{a}_{U,t}]$.
- *Reward*: The reward of each user is used to capture the benefit of a selected action. To maximize the similarities of all users in (6), the reward of each user i at time slot t is $w\left(\boldsymbol{a}_t \left| \boldsymbol{\xi}_t \right.\right) = \sum_{i \in \mathcal{U}} M\left(\mathcal{S}_{i,t}'\right)$ where $M\left(\mathcal{S}_{i,t}'\right)$ depends on $\boldsymbol{a}_{i,t}$. To calculate $M\left(\mathcal{S}_{i,t}'\right)$, we first use an attention network [11] to calculate the importance of each $\boldsymbol{s}_i^j$ in (1). Then, we can use $\boldsymbol{a}_{i,t}$ and constraint (6d) to determine $\mathcal{S}_{i,t}'$. Finally, we can obtain $M\left(\mathcal{S}_{i,t}'\right)$.
- *Q function*: A Q function $Q\left(\boldsymbol{\xi}_{i,t}, \boldsymbol{a}_{i,t}\right)$ of each user i is used to estimate the expected future reward under a given each state $\boldsymbol{\xi}_{i,t}$ and a selected action $\boldsymbol{a}_{i,t}$. The input of a Q function is a state while the output of the Q function is a vector of expected rewards resulting from each action at current state. We use a DNN to approximate the value function of each user i, which is parametrized by $\boldsymbol{\vartheta}_i$. After the training, each user can use Q function to find the optimal RB allocation and partial semantic information so as to maximize the similarity of all users.

B. Training of VD based DQN

The training of VD based DQN is jointly performed by the BS and the users. Next, we introduce their training process separately.

1) Training Process at Each User: Each user i needs to update its Q function using its collected data samples. The update of its Q function is

$$\boldsymbol{\vartheta}_i \leftarrow \boldsymbol{\vartheta}_i + 2\lambda_t \sum_{t=1}^{T} A\left(\boldsymbol{a}_t, \boldsymbol{\xi}_t\right) \nabla_{\boldsymbol{\vartheta}_i} Q\left(\boldsymbol{\xi}_{i,t}, \boldsymbol{a}_{i,t}\right), \tag{7}$$

where $A\left(\boldsymbol{a}_t, \boldsymbol{\xi}_t\right) = w\left(\boldsymbol{a}_t \left| \boldsymbol{\xi}_t \right.\right) + \gamma \max_{\boldsymbol{a}_i'} \sum_{i=1}^{U} Q\left(\boldsymbol{\xi}_{i,t+1}, \boldsymbol{a}_i'\right) - \sum_{i=1}^{U} Q\left(\boldsymbol{\xi}_{i,t}, \boldsymbol{a}_{i,t}\right)$, $\nabla_{\boldsymbol{\vartheta}_i} Q\left(\boldsymbol{\xi}_{i,t}, \boldsymbol{a}_{i,t}\right)$ is the gradient of $Q\left(\boldsymbol{\xi}_{i,t}, \boldsymbol{a}_{i,t}\right)$ with respect to $\boldsymbol{\vartheta}_i$, and λ_t is the step size.

2) Training Process at the BS: From (7), we can see that each user i updating its Q function needs the reward and the values of the Q functions from other users. Therefore, at each

Algorithm 1 VD based DQN for Solving Problem (6)

1: Initialize the DQN model $\boldsymbol{\vartheta}_i$ of each user i.
2: **for** each iteration **do**
3: **for** each user i **do**
4: **for** each time slot t **do**
5: Observe and record the location and energy level $\boldsymbol{\xi}_{i,t}$.
6: Select an action according to a ε-greedy scheme.
7: Calculate the reward $M(\boldsymbol{a}_{i,t})$ and Q function values $Q(\boldsymbol{\xi}_{i,t}, \boldsymbol{a}_{i,t})$ and $\max_{\boldsymbol{a}_i'} Q(\boldsymbol{\xi}_{i,t+1}, \boldsymbol{a}_i')$.
8: **end for**
9: Transmit the rewards and Q function values to the BS.
10: **end for**
11: The BS calculates $A(\boldsymbol{a}_t, \boldsymbol{\xi}_t)$ in (7) using the received rewards and Q function values.
12: The BS sends $A(\boldsymbol{a}_t, \boldsymbol{\xi}_t)$ to all users.
13: **for** each user i **do**
14: Use $A(\boldsymbol{a}_t, \boldsymbol{\xi}_t)$ to update $\boldsymbol{\vartheta}_i$ based on (7).
15: **end for**
16: **end for**

training step t, each user will first transmit its reward and the Q function values to the BS. Then, the BS will calculate $w(\boldsymbol{a}_t | \boldsymbol{\xi}_t)$, $\max_{\boldsymbol{a}_i'} \sum_{i=1}^{U} Q(\boldsymbol{\xi}_{i,t+1}, \boldsymbol{a}_i')$, and $\sum_{i=1}^{U} Q(\boldsymbol{\xi}_{i,t}, \boldsymbol{a}_{i,t})$ and transmit them to all users.

The entire offline training process of the VD based DQN is summarized in Algorithm 1. In particular, each user first use its DQN to select actions and calculate rewards. Then, it transmits the rewards and Q function values to the BS which aggregates the rewards and Q function values of all users and transmits them back to all users. Finally, each user can use the received rewards and Q function values to update its DQN parameters. In consequence, training VD based DQN only needs each user to share its rewards and Q function values. Here, we need to note that we use a mini batch method to update the VD based DQN and hence, the code from line 4 to line 8 must be implemented several times per iteration.

IV. Simulation Results and Analysis

For simulations, we consider a circular network area having a radius $r = 500$ m. $U = 10$ uniformly distributed users send semantic information to a BS that is at the center of the network area. The BS can allocate a total of $R = 5$ RBs to all users and thus only 5 users can transmit their semantic information the BS at each time slot. The bandwidth B of each RB is 2 MHz, the transmit power P of each device is 1 W, and $N_0 = -174$ dBm/Hz. For comparison purposes, we use independent DRL method in which each user uses a DQN to maximize its similarity without considering other users' actions. Our code is available at: https://github.com/wyy0206/ICASSP_semantic.

Fig. 1 shows the convergence of training the proposed VD based DQN and independent DRL. From this figure, we can see that the proposed method can improve the sum of all users' similarities by up to three-fold compared to the independent DRL. This is due to the fact that the proposed RL enables the users to jointly optimize their RB allocation and semantic information transmission schemes thus maximizing the sum of all users' similarities. In particular, in the proposed method, each user considers the sum of all users' similarities as a reward, and updates a DQN using the reward and Q function values of other users. In contrast, in the independent DRL, each user maximizes its own similarity and ignore the effects of other users.

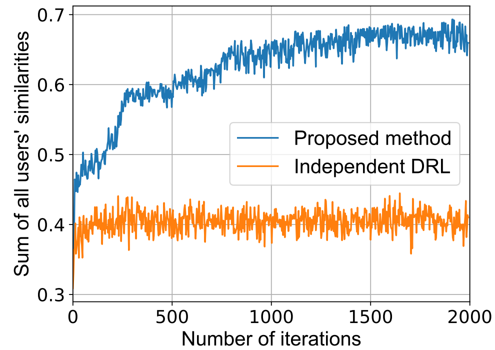

Fig. 1. Convergence of the proposed VD based DQN and independent RL.

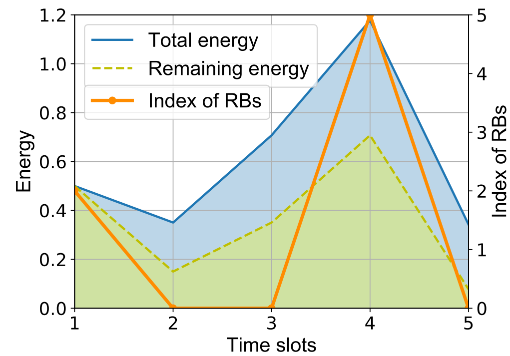

Fig. 2. The total energy, remaining energy, and RB allocation varies as time elapses. The index of RB implies the RB allocated to the user.

In Fig. 2, we show how the total energy, remaining energy, and RB allocation of a user change as time elapses. In Fig. 2, the gap between the total energy and remaining energy is the harvested energy. From Fig. 2, we can see that the user transmit semantic information at time slot 1 and does not transmit semantic information at time slots 2 and 3. This is because the user spends most of its energy for semantic information transmission at time slot 1 and it may not have enough energy to transmit information at time slots 2 and 3. In Fig. 2, we can also see that the total energy increases as the time slot changes from 2 to 4. This is due to the fact that, at time slots 2 and 3, the user does not use energy for information transmission and harvest.

V. Conclusion

In this paper, we have studied the optimization of semantic communication over energy harvesting networks. We have formulated an optimization problem that aims to maximize the similarities of all users while meeting the energy consumption constraint via optimizing the RB allocation scheme and the amount of semantic information to be transmitted for each user. To solve the proposed optimization problem, a novel collaborative and distributed RL algorithm was designed. In particular, the proposed algorithm enables the users to optimize their RB allocation and semantic information transmission schemes using their local information. Simulation results have shown that the proposed method significantly outperforms independent DRL.

REFERENCES

[1] Y. Shi, G.and Xiao, Y. Li, and X. Xie, “From semantic communication to semantic-aware networking: Model, architecture, and open problems,” *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44–50, Aug. 2021.

[2] E. C. Strinati and S. Barbarossa, “6G networks: Beyond shannon towards semantic and goal-oriented communications,” *Computer Networks*, vol. 190, pp. 107930, May 2021.

[3] Z. Yang, M. Chen, K. K. Wong, H. V. Poor, and S. Cui, “Federated learning for 6G: Applications, challenges, and opportunities,” Available Online*: https://arxiv.org/abs/2101.01338*, 2021.

[4] Z. Weng and Z. Qin, “Semantic communication systems for speech transmission,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2434–2444, Aug. 2021.

[5] H. Xie, Z. Qin, and G. Y. Li, “Task-oriented multi-user semantic communications for multimodal data,” Available Online*: https://arxiv.org/abs/2108.07357*, 2021.

[6] Y. Luan, L. He, M. Ostendorf, and H. Hajishirzi, “Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction,” in *Proc. Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, Oct. 2018.

[7] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, “A joint learning and communications framework for federated learning over wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269–283, Jan. 2021.

[8] R. Koncel-Kedziorski, D. Bekal, Y. Luan, M. Lapata, and H. Hajishirzi, “Text generation from knowledge graphs with graph transformers,” in *Proc. Conference of the North American Chapter of the Association for Computational Linguistics*, Minneapolis, MN, USA, June 2019.

[9] S. Prabhakaran, “Cosine similarity-understanding the math and how it works (with python codes),” https://www.machinelearningplus.com/nlp/cosine-similarity/, Sept. 2021.

[10] Y. Hu, M. Chen, W. Saad, H. V. Poor, and S. Cui, “Distributed multi-agent meta learning for trajectory design in wireless drone networks,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 3177–3192, Oct. 2021.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, Long Beach, CA, USA, Dec. 2017.