Resource Allocation for Semantic-Aware Networks
Lei Yan, Zhijin Qin, Rui Zhang, Yongzhao Li, Geoffrey Ye Li

Abstract—Semantic communications have shown its great potential to improve the transmission reliability, especially in low signal-to-noise regime. However, the resource allocation for semantic-aware networks still remains unexplored, which is a critical issue in guaranteeing the transmission reliability of semantic symbols and the communication efficiency of users. To fill this gap, we investigate the spectral efficiency in the semantic domain and rethink the semantic-aware resource allocation issue. Specifically, the semantic spectral efficiency (S-SE) is defined for the first time, and is used to optimize resource allocation in semantic-aware networks in terms of channel assignment and the number of transmitted semantic symbols. Additionally, for fair comparison of semantic and conventional communication systems, a transform method is developed to convert the conventional bit-based spectral efficiency to the S-SE. Simulation results demonstrate the validity and feasibility of the proposed semantic-aware resource allocation model, as well as the superiority of semantic communications in terms of the S-SE.

Index Terms—Semantic communications, semantic spectral efficiency, resource allocation.

I. INTRODUCTION

With growing wireless applications and increasing data traffic volumes, wireless communications are facing the bottleneck of spectrum scarcity, which motivates a paradigm shift from conventional to semantic communications [1], [2]. By focusing on transmitting the meaning of the source information, semantic communications have shown the great potential to reduce the network traffic and thus alleviate spectrum shortage. Particularly, different semantic communication systems have been designed for different types of sources, including text [3], [4], image [5], [6], and speech [7], yielding a significant improvement in the transmission reliability of semantic symbols. Nevertheless, as for improving the communication efficiency of users while guaranteeing the transmission reliability, it is vital to investigate the resource allocation issue for semantic-aware networks [8].

In wireless communications, how to measure the information content as well as the spectral efficiency (SE) is fundamental to the study of resource allocation. In conventional communications, bits are used to quantify them. However, this manner is not applicable in semantic communications, because bits are used to quantify the information content as well as the amount of it in a sentence based on the logical probability. Carnap [9] first attempted to define the semantic information and the amount of it in a sentence based on the logical probability. On this basis, the semantic channel capacity was derived in [10] for the discrete memoryless channel, revealing the existence of semantic coding strategy for reliable communications. Furthermore, semantic coding was investigated in [11], and the fundamental limits of semantic transmission as well as semantic compression were studied. However, the aforementioned works are based on abstract models with limited possibility for implementation, and failed to quantify the SE in the semantic domain.

Although a complete theory or a well-developed mathematical model for semantic communications is still missing, the success of semantic communication system design with the aid of deep learning (DL) makes it possible to define a calculable SE in the semantic domain. Particularly, the DL-enabled semantic communication system (DeepSC) was proposed in [3], and has been extended to several variants [4], [12]. They can effectively extract the semantic information from text and successfully deliver the desired meaning to the receiver. In this article, we implement the DeepSC as an example to explore the SE issue in the semantic domain and investigate the resource allocation problem in such a semantic-aware network. The main novelties and contributions can be summarized as follows:

- A novel resource allocation model is proposed for semantic-aware networks. Specifically, the semantic spectral efficiency (S-SE) is first defined to measure the communication efficiency from the semantic perspective. Then a new formulation is proposed and solved to maximize the S-SE in terms of channel assignment and the number of transmitted semantic symbols.
- To make a fair comparison between semantic and conventional communication systems, a transform method is developed to convert the bit-based SE to the S-SE.
- Simulation results verify the effectiveness of the proposed resource allocation model, as well as the superiority of semantic communication systems in terms of the S-SE.

The rest of this article is organized as follows. Section II introduces the system model. The proposed semantic-aware resource allocation is formulated and solved in Section III. Section IV proposes a transform method for fair comparisons of semantic and conventional communication systems, and presents the simulation results. Section V concludes the article.

Notation: $\mathbb{R}^{n \times m}$ represents the set of real matrices of size $n \times m$. Bold-font variables represent matrices and vectors. $x \sim \mathcal{CN} (\mu, \sigma^2)$ means $x$ follows a circularly-symmetric complex Gaussian distribution with mean $\mu$ and covariance $\sigma^2$.

II. SYSTEM MODEL

We consider a cellular network, which consists of a base station (BS) and a set of users denoted by
In conventional communications, spectral efficiency is measured in bits per second per Hertz (bits/s/Hz), which can effectively measure the transmission rate of bit sequences but cannot be used to measure the transmission rate of semantic information. This is because the bit sequences are produced according to the statistical knowledge of the source, which are less relevant to the underlying meaning of the source. Thus new performance metrics need to be investigated at the semantic level.

In this section, the S-SE is first defined as a new metric for the semantic-aware network. Then the semantic-aware resource allocation is formulated as a S-SE maximization problem in terms of channel assignment and the number of transmitted semantic symbols. Finally, the optimal solution of the optimization problem is obtained.

### A. Semantic Spectral Efficiency

In conventional communications, spectral efficiency is measured in bits per second per Hertz (bits/s/Hz), which can effectively measure the transmission rate of semantic information per second, and is measured in suts/s.

Semantic spectral efficiency (S-SE) refers to the rate at which semantic information can be successfully transmitted over a unit of bandwidth, and is measured in suts/s. The expression of S-SE is given by:

$$S_{SE} = \frac{1}{MN} \sum_{n=1}^{N} \sum_{m=1}^{M} \log_2(1 + \gamma_{n,m}),$$

where $p_n$ is the transmit power of the $n$-th user, $g_n$ is the large-scale channel gain of the $n$-th user including path loss and shadowing, $h_{n,m} \sim CN(0,1)$ is the Rayleigh fading coefficient for the $n$-th user transmitting over the $m$-th channel, and $N_0$ is the noise power spectral density.

**C. DeepSC Receiver**

At the BS, the signal from the $n$-th user can be denoted as $Y_n = \sqrt{g_n h_{n,m}} X_n + z$ where $z \sim CN(0,1)$ is the Gaussian noise. Then the signal will be decoded by the channel decoder and the semantic decoder to estimate the sentence $s_n$.

In order to evaluate the performance of semantic communications for text transmission, we adopt the semantic similarity as the performance metric, which is given by

$$\xi = \frac{\mathbf{B}(\hat{s})^T \mathbf{B}(\hat{s})}{||\mathbf{B}(\hat{s})||^2},$$

where $\mathbf{B}(\cdot)$ represents Sentence-Bidirectional Encoder Representations from Transformers (BERT) model [13], which achieves a significant improvement over state-of-the-art sentence embedding methods. Compared with other semantic metrics, such as bilingual evaluation understudy (BLEU) [14], BERT-level similarity can measure the distance of semantic information between two sentences more precisely. The output range of this parameter falls between 0 and 1. Here, $\xi = 1$ means that two sentences has highest similarity, and $\xi = 0$ means no similarity between them.

### III. SEMANTIC-AWARE RESOURCE ALLOCATION

In this section, the S-SE is first defined as a new metric for the semantic-aware network. Then the semantic-aware resource allocation is formulated as a S-SE maximization problem in terms of channel assignment and the number of transmitted semantic symbols. Finally, the optimal solution of the optimization problem is obtained.

#### A. Semantic Spectral Efficiency

In conventional communications, spectral efficiency is measured in bits per second per Hertz (bits/s/Hz), which can effectively measure the transmission rate of bit sequences but cannot be used to measure the transmission rate of semantic information. This is because the bit sequences are produced according to the statistical knowledge of the source, which are less relevant to the underlying meaning of the source. Thus new performance metrics need to be investigated at the semantic level.

For the sake of clarity, we first assume that semantic information can be measured by the semantic unit (suts), which represents the basic unit of semantic information. Based on this, two crucial semantic-based performance metrics can be defined:

- **Semantic transmission rate (S-R)** refers to the effectively transmitted semantic information per second, and is measured in suts/s.
- **Semantic spectral efficiency (S-SE)** refers to the rate at which semantic information can be successfully transmitted over a unit of bandwidth, and is measured in suts/s.

Then the expressions of S-R and S-SE are derived respectively in the following. Taking $D = \{s_j = [w_{j,1}, w_{j,2}, \ldots, w_{j,L_j}]\}_{j=1}^D$ with size $D$ as the text
dataset, where \( s_j \) is the \( j \)-th sentence with length \( L_j \) in the set and \( w_{j,t} \) is the \( t \)-th word, we denote the amount of semantic information of \( s_j \) by \( I_j \). With \( p(s_j) \) representing the occurrence probability of \( s_j \), the expected amount of semantic information per sentence can be expressed as \( I = \sum_{j=1}^{D} I_j p(s_j) \), which corresponds to an expected number of words per sentence as \( L = \sum_{j=1}^{D} L_j p(s_j) \). Note that we focus on the network performance for the long-term text transmission, rather than the transmission of individual sentences, so the expected values \( I \) and \( L \), instead of the random values, should be taken to obtain the representations of S-R and S-SE. Hence, at the \( n \)-th user, we consider that \( k_n L \) semantic symbols carries the amount of semantic information of \( I \). Thus, the average amount of semantic information per semantic symbol can be expressed as \( I/L \). Moreover, since the symbol rate is equal to the channel bandwidth for passband transmission, the total semantic information transmitted over the channel with the bandwidth \( W \) is \( WI/L(k_n) \). Thus the S-R of the \( n \)-th user over the \( m \)-th channel can be expressed as

\[
\Gamma_{n,m} = \frac{WI}{k_n L} \xi_{n,m} \tag{4}
\]

where \( \xi_{n,m} \) is the semantic similarity of the \( n \)-th user over the \( m \)-th channel. Note that \( \xi_{n,m} \) relies on the neural network structure of DeepSC and channel conditions, which can be expressed as a function of \( k_n \) and \( \gamma_{n,m} \), i.e., \( \xi_{n,m} = f(k_n, \gamma_{n,m}) \).

Based on the expression of the S-R, the S-SE of the \( n \)-th user over the \( m \)-th channel can be expressed as

\[
\Phi_{n,m} = \frac{\Gamma_{n,m}}{W} = \frac{I}{k_n L} \xi_{n,m} \tag{5}
\]

### B. Problem Formulation

In this part, a semantic-aware resource allocation model is proposed to maximize the S-SE of the network. By denoting \( \Phi \) as the S-SE of the network, we have

\[
\Phi = \sum_{n=1}^{N} \sum_{m=1}^{M} \xi_{n,m} \frac{I}{k_n L} \tag{6}
\]

As for the optimization variables, we consider not only the common optimization variables involved in the conventional resource allocation model, i.e., the channel assignment vector, but also the average number of transmitted semantic symbols for each word \( k_n \), which is closely related to the system performance and should be optimized to take full advantage of semantic communications. Consequently, the constraint about \( k_n \) should be considered as well.

According to the above analysis, the S-SE maximization problem can be formulated as

\[
\text{(P0)} \quad \max_{\alpha_n,k_n} \Phi \tag{7}
\]

\[\text{s.t.} \quad C_1 : \alpha_{n,m} \in \{0, 1\}, \forall n \in \mathcal{N}, \forall m \in \mathcal{M} \tag{7a}\]

\[C_2 : \sum_{n=1}^{N} \alpha_{n,m} \leq 1, \forall m \in \mathcal{M} \tag{7b}\]

\[C_3 : \sum_{m=1}^{M} \alpha_{n,m} \leq 1, \forall n \in \mathcal{N} \tag{7c}\]

\[C_4 : 1 \leq k_n \leq K \tag{7d}\]

\[C_5 : \xi_{n,m} \geq \xi_{th} \tag{7e}\]

\[C_6 : \Phi_{n,m} \geq \Phi_{th} \tag{7f}\]

where \( C_1, C_2, \) and \( C_3 \) are channel assignment constraints, \( C_4 \) specifies the permitted range of the average number of semantic symbols per word with \( K \) representing the maximum value, \( C_5 \) reflects the minimum required semantic similarity \( \xi_{th} \), and \( C_6 \) restricts the minimum S-SE of users by \( \Phi_{th} \).

### C. The Optimal Solution

To solve (P0), two challenges should be addressed. One is how to deal with the term \( I/L \) in the objective function, and the other is how to obtain the semantic similarity \( \xi_{n,m} \), which is closely related to the objective function \( \Phi \), \( C_5 \), and \( C_6 \).

Regarding the first challenge, we note that the term \( I/L \) depends on the type of source. According to the previous analysis in Section III-A, this term is a constant for a particular type of source, which will not affect the resource optimization. Consequently, we can omit this term when solving (P0). Thus the optimization problem (P0) can be rewritten as

\[
\text{(P1)} \quad \max_{\alpha_n,k_n} \Phi = \sum_{n=1}^{N} \sum_{m=1}^{M} \alpha_{n,m} \xi_{n,m} \frac{I}{k_n L} \tag{8}
\]

\[\text{s.t.} \quad C_1, C_2, C_3, C_4, C_5, C_6 \tag{8b}\]

Regarding the second challenge, since \( \xi_{n,m} \) is dependent of the specific semantic communication system and the physical channel conditions, we run the DeepSC model under different conditions to obtain a table reflecting the mapping between \( \xi_{n,m} \) and \( (k_n, \gamma_{n,m}) \), as shown in Fig. 2.

After addressing the two challenges, (P0) can be solved. Specifically, due to the orthogonality of different cellular links, (P1) can be decoupled into the following two equivalent independent optimization problems:

\[
\text{(P2)} \quad \max_{k_n} \Phi_{n,m} \tag{9}
\]

\[\text{s.t.} \quad C_4, C_5, C_6 \tag{9b}\]

and

\[
\text{(P3)} \quad \max_{\alpha_n} \Phi_{n,m}^{\max} \tag{10}
\]

\[\text{s.t.} \quad C_1, C_2, C_3 \tag{10b}\]

where \( \Phi_{n,m} = \xi_{n,m}/k_n \) and \( \Phi_{n,m}^{\max} \) represents the maximum \( \Phi_{n,m} \) with respect to \( k_n \). (P2) targets on obtaining \( \Phi_{n,m}^{\max} \) for all users over all candidate channels. Since \( \xi_{n,m} \) in \( C_5 \) and \( C_6 \) can only be obtained by the look-up table method, the exhausted searching method is adopted to solve (P2). Moreover,
IV. SIMULATION RESULTS AND COMPARISONS

In order to evaluate the performance of the proposed semantic-aware resource allocation scheme comprehensively, we conduct the following verifications in the simulation:

1) Compare the proposed resource allocation model against the conventional one to verify the proposed model in semantic-aware networks.
2) Compare the S-SE of semantic and conventional communication systems to show the superiority of semantic communications.

Considering that conventional communication systems are usually assessed in the bit domain, we first develop a transform method to convert the typical SE to the S-SE to compare the two communication paradigms fairly. On this basis, simulation results are presented and analysed.

A. The Transform Method for Fair Comparisons

In conventional communications, each letter in a word is mapped into bits through source encoder. From the semantic perspective, each bit can be regarded as a semantic symbol, although it may carry less semantic information than the semantic symbol of DeepSC. Similar to the definitions in Section III-A, the equivalent S-R can be expressed as

\[ \Gamma_{n,m}' = C_{n,m} \frac{I}{\mu L} \xi_{n,m} \]

where \( C_{n,m} \) is the transmission rate of the \( n \)-th user over the \( m \)-th channel, measured in bits/s. Let us define \( \mu \) as the transforming factor related to the source coding scheme, representing the average number of bits per word, and measured in bits/word. Specifically, assuming that a word includes five letters on average, and ASCII code is adopted to encode each letter, we have \( \mu = 40 \) bits/word. Moreover, assuming that there is no bit error in conventional communications, thus \( \xi_{n,m} \) is equal to 1. By denoting \( R_{n,m} = C_{n,m} / W \) as the SE, the equivalent S-SE can be given by

\[ \Phi_{n,m}' = R_{n,m} \frac{I}{\mu L}. \]

By doing so, we can replace the objective function of the conventional resource allocation scheme from SE to S-SE and make it possible to compare with semantic communication systems.

B. Benchmarks

Considering that the proposed semantic-aware resource allocation scheme is based on a specific semantic communication system, i.e., DeepSC, we consider the following three benchmarks:

- **Ideal system**: Shannon limit can be achieved with no bit errors.
- **4G system**: According to the measured SNR, the BS obtains the channel quality indicator (CQI) [16], based on which the achievable SE can be obtained according to Table 7.2.3-1 in 3GPP TS 36.213.
- **5G system**: Similar to 4G, the BS obtains CQI based on the measured SNR [17], and then obtains the achievable SE according to Table 5.2.2.1-2 in 3GPP TS 38.214.

Table 1: Simulation parameters.

| Parameter                  | Value               |
|----------------------------|---------------------|
| Number of users, \( N \)   | 5                   |
| Number of channels, \( M \)| 5                   |
| Channel bandwidth, \( W \) | 180 KHz             |
| Noise power spectral density, \( N_0 \) | -174 dBm/Hz         |
| Path loss model            | 128.1+37.6lg[\( d(km) \)] dB |
| Shadow effect factor       | 6 dB                |
| Transmit power, \( P_0 \)  | 10 dBm              |
| Maximum number of symbols per word, \( K \) | 20 symbols/word     |
| Semantic similarity threshold, \( \xi_{th} \) | 0.9                 |
| S-SE threshold, \( \Phi_{th} \) | 0.025(I/L) suts/s/Hz |
| Transforming factor, \( \mu \) | 40 bits/word        |

For the above three benchmarks, the S-SE maximization problem can be formulated as

\[
\begin{aligned}
\text{(P4)} \quad & \max_{\alpha_{n,m}} \sum_{n=1}^{N} \sum_{m=1}^{M} \alpha_{n,m} \Phi_{n,m}' \\
\text{s.t.} \quad & C_1, C_2, C_3 \\
& C_7: \Phi_{n,m}' \geq \Phi_{th},
\end{aligned}
\]

where \( \Phi_{n,m}' \) is the S-SE of the \( n \)-th user over the \( m \)-th channel in system \( \Delta \), \( \Delta \in \{\text{Ideal, 4G, 5G}\}. \) These optimization problems can be solved by the method introduced in Section III-C.

C. Simulation Results

In our simulation, a circular network with radius \( r = 500 \) m is considered where \( N \) users are distributed uniformly. Unless specified otherwise, the relevant parameters are listed in Table I.

We first examine the conventional resource allocation model in semantic-aware networks. In this simulation, the optimal channel assignment results of the conventional model in the ideal system is applied to the network, along with different values of \( k_n \). Then the obtained S-SE is compared with that of the proposed model. As shown in Fig. 3, the S-SE of the conventional model is smaller than that of the proposed model regardless of the value of \( k_n \), which implies that the conventional model is not suitable in semantic-aware networks. In addition, the S-SE of the conventional model with \( k_n = 3 \) is equal to 0 because the semantic similarity is less than the threshold in this case.

In the following, we compare the S-SE of different communication systems with the corresponding resource allocation model.
model. Fig. 4(a) shows the S-SE of different systems versus the number of channels. When $M$ is increased from 1 to 5, the S-SE of all systems increases rapidly because more users are served. Then when $M$ is increased from 5 to 10, the S-SE grows gently instead of remaining stable because more channels are available and users can choose the channel with higher SNR. Moreover, the semantic communication system outperforms all conventional communication systems.

Fig. 4(b) illustrates the S-SE versus the transmit power. As $p_n$ increases, the S-SE of the ideal system increases rapidly, while that of the semantic communication system, 4G system, and 5G system increase first and then tend to be a constant, implying that all practical systems have an upper bound for the S-SE with increasing SNR. Moreover, the semantic communication system shows a larger upper bound than 4G and 5G due to its stronger ability in compressing data.

Fig. 4(c) shows the S-SE versus the transforming factor. From this figure, the S-SE of the semantic communication system remains stable since the transforming factor is irrelevant to semantic communications. For the conventional systems, the S-SE decreases with $\mu$ increasing because the S-SE is the ratio of the SE and $\mu$, and the maximum SE is a fixed value with different $\mu$. Additionally, the semantic communication system yields better performance than both 4G and 5G when $\mu$ is larger than 21 bits/word. Nevertheless, when $\mu$ is smaller than approximately 30 bits/word, i.e., a word can be encoded to less than 30 bits, the semantic communication system performs worse than the ideal system. This figure demonstrates that whether semantic communication systems outperform conventional ones to a great extent depends on the source coding scheme adopted in conventional systems.

V. CONCLUSIONS

In this letter, we have studied the SE issue in semantic domain and explored the resource allocation for semantic-aware networks. Specifically, S-R and S-SE have been defined first to make it possible to measure the communication efficiency of the semantic communication system based on the DeepSC model [3]. Aiming at maximizing the S-SE of the network, the semantic-aware resource allocation has been formulated as an optimization problem, and the optimal solution has been obtained. Extensive simulation has been conducted to evaluate the performance of the proposed scheme. An insightful conclusion is that, for text transmission, semantic communication systems achieve a higher S-SE than both 4G and 5G systems when a word is mapped to more than 21 bits on average through conventional source coding techniques, and if the required bits for encoding a word is increased to more than 30 bits with 10 dBm transmit power, semantic communication systems even outperforms the ideal system.

REFERENCES

[1] W. Tong and G. Y. Li, “Nine challenges in artificial intelligence and wireless communications for 6G,” to appear in IEEE Wireless Commun., Sept. 2021.
[2] Z. Qin, X. Tao, J. Lu, and G. Y. Li, “Semantic communications: Principles and challenges,” Dec. 2021, arXiv: 2201.01389.
[3] H. Xie, Z. Qin, G. Y. Li, and B. H. Juang, “Deep learning enabled semantic communication systems,” IEEE Trans. Signal Process., vol. 69, pp. 2663–2675, Apr. 2021.
[4] M. Sana and E. C. Strinati, “Learning semantics: An opportunity for effective 6G communications,” Oct. 2021, arXiv: 2110.08049.
[5] C.-H. Lee, J.-W. Lin, P.-H. Chen, and Y.-C. Chang, “Deep learning-constructed joint transmission-recognition for internet of things,” IEEE Access, vol. 7, pp. 76547–76561, Jun. 2019.
[6] E. Bourtsoulatze, D. Burth Kurka, and D. Gündüz, “Deep joint source-channel coding for wireless image transmission,” IEEE Trans. Cogn. Commun. Netw., vol. 5, no. 3, pp. 567–579, May 2019.
[7] Z. Weng and Z. Qin, “Semantic communication systems for speech transmission,” IEEE J. Sel. Areas Commun., vol. 39, no. 8, pp. 2434–2444, Jun. 2021.
[8] M. Kountouris and N. Pappas, “Semantics-empowered communication for networked intelligent systems,” IEEE Commun. Mag., vol. 59, no. 6, pp. 96–102, Jul. 2021.
[9] R. Carnap and Y. Bar-Hillel, “An outline of a theory of semantic information,” Res. Lab. Electronics, Massachusetts Inst. Technol., Cambridge MA, RLE Tech. Rep. 247, Oct. 1952.
[10] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, “Towards a theory of semantic communication,” in Proc. IEEE Netw. Sci. Workshop, West Point, NY, USA, Jun. 2011, pp. 110–117.
[11] F. Willems and T. Kalker, “Semantic compaction, transmission, and compression codes,” in Proc. Int. Symp. Inform. Theory (ISIT), Adelaide, SA, Australia, Sept. 2005, pp. 214–218.
[12] H. Xie, Z. Qin, and G. Y. Li, “Task-oriented multi-user semantic communications for VQA task,” IEEE Wireless Commun. Lett., Dec. 2021, early access.
[13] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using siamese BERT-networks,” in Proc. Emp. Methods in Natur. Lang. Process. (EMNLP), Nov. 2019.
[14] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” in Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL), Philadelphia, Pennsylvania, USA, Jul. 2002, pp. 311–318.
[15] S. W. Kuhn, “The hungarian method for the assignment problem,” Naval Res. Logist. Quart., vol. 2, pp. 83–85, Mar. 1955.
[16] A. Ghosh and R. Ratasuk, Essentials of LTE and LTE-A, Cambridge, U.K.: Cambridge Univ. Press, 2011.
[17] E. Chu, J. Yoon, and B. C. Jung, “A novel link-to-system mapping technique based on machine learning for 5G/6G wireless networks,” Sensors, vol. 19, no. 5, Mar. 2019.