Guest Editorial
Special Issue on Beyond Transmitting Bits: Context, Semantics, and Task-Oriented Communications

Deniz Gündüz, *Fellow, IEEE*, Zhijin Qin, *Senior Member, IEEE*, Inaki Estella Aguerri, Harpreet S. Dhillon, *Senior Member, IEEE*, Zhaohui Yang, Aylin Yener, *Fellow, IEEE*, Kai Kit Wong, *Fellow, IEEE*, and Chan-Byoung Chae, *Fellow, IEEE*

It is our pleasure to share with you this Special Issue, which brings together a diverse set of articles dealing with various aspects of semantic and goal-oriented communications, providing a snapshot of research activities in this highly active research area. Wireless communications and networking research has traditionally focused on improving the capacity and throughput of the underlying wireless network. However, recent explosion in data-driven machine learning applications and their reliance on huge datasets collected by edge devices have raised legitimate concerns that the increasing data traffic might soon overwhelm the capacity of current networks despite ongoing efforts to increase their capacity and efficiency. Also, most of the edge intelligence applications impose stringent delay constraints, which cannot be met by naive forwarding of data samples for processing at the receiver end. This made it obvious to researchers in both academia and industry that it is essential to analyze the "value" or "relevance" of collected data, and filter and prioritize the delivery of data based on its value/relevance as well as the wireless channel and network conditions. In this context, data value will be closely connected to the underlying signals and processes that generate the data, e.g., text, image, video, or sensor data, and what the receiver intends to do with the received data. This subjectivity of data value makes semantic and goal-oriented communication a rather elusive research topic, which has led to both an increasingly rich and active area of investigation, but also a controversial one, mainly due to the lack of clear and widely agreed-upon definitions of some of the core concepts and formulations. Despite these disagreements, there is almost unanimous consensus on the importance and potential impact of this line of investigation for the design of future communication systems and networks.

Deniz Gündüz is with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: d.gunduz@imperial.ac.uk).

Zhijin Qin is with the Department of Electronic Engineering, Tsinghua University, Beijing 100190, China (e-mail: qinzhijin@tsinghua.edu.cn).

Inaki Estella Aguerri is with Amazon, 28045 Barcelona, Spain (e-mail: inaki.estella@gmail.com).

Harpreet S. Dhillon is with the Wireless@Virginia Tech, Bradley Department of ECE, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: hdhillon@vt.edu).

Zhaohui Yang is with Zhejiang Lab, Hangzhou 311121, China, also with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China, and also with the Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking (IPCAN), Hangzhou, Zhejiang 310007, China (e-mail: yang_zhaohui@zju.edu.cn).

Aylin Yener is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: yener@ece.osu.edu).

Kai Kit Wong is with the Department of Electronic and Electrical Engineering, University College London, WC1E 6BT London, U.K. (e-mail: kai-kit.wong@ucl.ac.uk).

Chan-Byoung Chae is with the School of Integrated Technology, Yonsei University, Seoul 03722, South Korea (e-mail: cbchae@yonsei.ac.kr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2022.3221853.

Digital Object Identifier 10.1109/JSAC.2022.3221853

The goal of this Special Issue was to encourage discussion among researchers with different perspectives by attracting and publishing creative approaches for the design of modern semantic and goal-oriented communication networks. We believe that we have achieved this goal as our call-for-papers received a strong response from the community: we received a total of 55 submissions, 17 of which have been accepted for publication. The high number of submissions is a clear indication of the growing interest in this area of research, and the diversity and the quality of the articles selected for publication is a testament to the rapid progress made in the last several years.

In the tutorial article written by the guest editorial team [A1], which has the same title as this Special Issue, we have presented an overview of the main approaches and challenges to research in this area, but also tried to provide a general and inclusive framework to understand ongoing research efforts and future challenges in semantic and goal-oriented communications. To emphasize the difficulty in providing a clear definition for semantics in communications, we have started the article with a historical perspective, highlighting that this quest goes back many decades, and many attempts have already been made by well-known researchers from different backgrounds. After a brief summary of studies about semantic entropy measures and knowledge graph approaches to extract semantic information, we mainly focused on the goal-oriented communication interpretation, where semantics is treated as a generic distortion measure used to quantify the quality of reconstruction at the receiver. However, we have shown that this goes beyond conventional rate-distortion problems focusing on additive single-letter distortion measures, and covers remote inference problems as well as rate-distortion with perception constraints. We have then treated the semantic communication over noisy channels as a joint source-channel coding problem, again with a generic fidelity measure. For both rate-distortion and joint source-channel coding approaches, we have especially focused on machine learning approaches to overcome the difficulty of lack of source-statistics or clear reconstruction measures through data-driven designs. We have tried to connect these practical approaches to information theoretical fundamentals. While these fundamental limits are based on various idealistic assumptions, the coding and communication techniques derived to achieve these limits often inspire model-based data-driven techniques for more practical scenarios. We have provided a comprehensive overview of one such tool, the information bottleneck, which arises both in various multi-user communication problems, and in remote inference problems. We have then introduced the pragmatic/effective communication problem, in which the communication takes place over a time-horizon, and involves a time-sensitive reconstruction measure. This leads to enabling not only high-fidelity but also timely reconstruction of signals over wireless channels, which also creates close connections between the semantic/goal-oriented communication problem and control over noisy channels.

Along the categories exposed in our tutorial article, the rest of the articles included in this Special Issue can be grouped into three areas: 1) semantic and goal-oriented compression and signal processing, 2) semantic and goal-oriented joint source-channel coding, and 3) pragmatic/effective communications. The contributions made by each of the articles are summarized next.

I. Semantic and Goal-Oriented Compression and Signal Processing

In [A2], Zou et al. consider a scenario in which a receiver has to execute a task from a quantized version of the information source of interest. The task is modeled as an optimization of a general goal function over a set of quantized parameters characterizing it. The authors resort to the high-resolution analysis and characterize the relationship between the function to optimize and other parameters for the scalar case and provide approximations for the multidimensional scenario. An algorithm is provided based on these insights to design practical quantizers. The numerical results show the advantages of the proposed algorithm over traditional approaches not accounting for quantization. The analysis also provides insights on the fundamental relationship between the goal function regularity properties and the difficulty of quantizing its parameters.

Given the significance of image transmission over networks, and the rich content of images that lend itself to a wide variety of inference tasks as well as reconstruction with various distortion and perception constraints, quite a few articles in this issue consider semantic image compression problems. In [A3], image semantic refers to different categories present in the image. Therefore, rather than focusing on classical image reconstruction metrics such as peak signal-to-noise ratio (PSNR) or structural similarity index measure (SSIM), the authors consider a combination of intersection of union (IoU), which measures the average segmentation precision, the object detection loss, which measures the accuracy of an object detection task on the reconstructed image, and a visual loss, which measures the mean-square error between the original and the reconstructed image in a feature space of a pre-trained model. Since these tasks are non-differentiable, and hence, cannot be backpropagated easily through a neural network architecture, the authors follow a reinforcement-learning-based approach, where the bit budget is allocated among the captured semantic concepts in a sequential manner. The experimental results demonstrate that the proposed approach can reconstruct visually pleasant and semantically consistent images in the low bit rate regime.

In [A4], Lei et al. focus on the progressive image compression problem taking into account the rate-distortion-classification-perception (RDCP) trade-off. In particular, the goal is for the reconstructed image to simultaneously achieve high visual quality and high performance in downstream inference tasks. The authors propose a latent-based rate-adaptive image compression algorithm using deep neural networks, which learns a structured latent representation that can provide both rate and context adaptation in resource-limited scenarios.

The next two articles consider the video compression problem. Designing semantic and goal-oriented video compression algorithms could have a significant potential impact since video content dominates the mobile data traffic today, and this is only expected to grow with the introduction of augmented reality (AR) and virtual reality (VR) applications at the wireless edge. In [A5], the authors propose a real-time video analytics system that leverages the rich contextual information of surveillance cameras to reduce the bandwidth consumption through semantic compression. The proposed method dynamically maintains the background image of the video at the edge server with minimal system overhead, and sends only highly confident region of interests to the cloud through adaptive weighting and encoding. With a lightweight experience-driven learning module, the proposed solution is able to achieve high offline inference accuracy even when network congestion occurs.

Another article focusing on goal-oriented video compression is [A6], in which Tandon et al. focus on video conferencing applications. The authors develop a low-bitrate codec by converting the speech to text at the transmitter and reconstructing the video at the receiver. Particularly, the Txt2Vid model is proposed by only transmitting a short video as the driving video and the text obtained by speech recognition at the transmitter. The video is reconstructed based on the user ID and the lips of the talking head area generated by the received text. The developed scheme could reduce the network traffic by two to three times without degrading the reconstruction quality in terms of subjective evaluation.

In [A7], Kalfa et al. propose a semantic information extraction framework for graph signals with intrinsic semantic noise, which consists of a raw semantic extractor, semantic fidelity control block, attribute tracking, signal smoothing and signal tracking. The proposed method is shown to provide reliable semantic outputs and enable efficient ways of identifying semantic innovation while filtering out semantic noise. The performance of the proposed methods is evaluated via simulations and real-world computer vision examples.

II. Semantic and Goal-Oriented Communication Over Wireless Channels

While the above articles mainly focus on the compression and signal processing aspects of semantic communication, the following group of articles take the channel noise into account, and design goal-oriented wireless communication techniques that are cognizant of the semantics of the underlying signals.

The first article in this group focuses mainly on the distributed inference problem over noisy channels. In [A8], Kim et al. consider the resource allocation problem for distributed classification over noisy channels, focusing on distributed boosting classifiers. Boosting classification algorithms make a final decision via a weighted vote among the outputs of base classifiers. However, when the base classifiers communicate their votes over wireless channels, channel noise can degrade the final classification accuracy. The authors consider the scenario in which the base classifiers communicate their votes over a shared wireless channel, and formulate a resource allocation problem by taking into account the importance of base classifiers and the available resource budget. This provides a good example of goal-oriented communications, where the goal is to improve the final classification accuracy, rather than to reliably recover individual transmitted bits. It is concluded that more system resources should be allocated to more important base classifiers to minimize noise-induced classification degradation.

In [A9], Mu et al. consider a multi-user communication problem, in which an access point transmits *semantic* and conventional *bit-based* information to two different users. The performance of semantic transmission is measured in terms of *semantic rate*, which depends upon *semantic similarity* whose closed-form characterization is not available. Authors overcome this challenge by approximating semantic similarity by a generalized logistic function through a data regression method. To enable the simultaneous transmission of the semantic and bit-based information, three multiple access schemes are considered, namely orthogonal multiple access (OMA), non-orthogonal multiple access (NOMA), and semi-NOMA. The performance of these schemes is characterized through the *SvB rate region*, which consists of all achievable semantic and bit rate-pairs, and the *power region*, which determines the minimum transmit power to achieve the target semantic and bit rates. The analysis demonstrates the superiority of semi-NOMA over both OMA and NOMA.

Several of the articles focus exclusively on wireless image transmission and propose semantic communication techniques that go beyond reconstruction of images with low PSNR. In [A10], Zhang et al. consider goal-oriented image transmission, where the underlying machine learning task to be performed at the receiver is not known by the transmitter, and the data distribution might be dynamic. The receiver has a limited feedback loop with the encoder, and the authors propose a neural network-based communication scheme consisting of two main parts, namely the semantic coding network and the data adaptation network. The semantic coding network includes an encoder for semantic data compression and a decoder for information recovery at the receiver, and learns how to extract and transmit the semantic information using a receiver-leading training process. The second part is a generative adversarial network deployed at the device that does data adaption using domain adaptation techniques from transfer learning to learn how to convert the observed data into a similar form of the empirical data that the semantic coding network can process without retraining. Numerical experiments validate that the proposed method can be adapted to dynamic datasets while keeping a flexible balance between the data recovery and task execution performances.

In [A11], Kang et al. consider the transmission of sensing images obtained by an unmanned aerial vehicle (UAV), and develop an energy-efficient task-oriented semantic image transmission framework with a scene graph based on subject-relation-object triplets. A personalized semantic encoder is designed based on user interests to meet personalized saliency requirements with minimal energy cost. A resource allocation scheme is also developed for a multi-user scenario in a dynamic channel environment.

In [A12], the authors introduce the concept of semantic slice-models (SeSMs) to enable semantic-aware wireless image transmission. In particular, semantic service quality is proposed as a new performance metric for semantic communication systems, and SeSM is used to provide adaptive reconstruction and perception quality. The performance of the proposed approach is studied for both compression and joint source-channel coding problems in the presence of feedback. The developed system is compared with other alternatives through numerical simulations.

Joint source-channel coding of video content is studied by the following two articles. In [A13], Wang et al. study deep joint source-channel coding over wireless channels with a focus on video transmission. The proposed methods exploit nonlinear transform and conditional coding architecture to adaptively extract semantic features across video frames, and transmit semantic feature domain representations over wireless channels via deep joint source-channel coding. A novel rate-adaptive transmission mechanism, which learns to allocate the limited channel bandwidth among video frames, is developed to maximize the end-to-end performance. Across standard video source test sequences and various communication scenarios, experiments show that the proposed approach can generally surpass traditional wireless video-coded transmission schemes.

In [A14], Jiang et al. explore wireless semantic communications for video conferencing. It is inspired by the fact that in video conferencing, the motion of the objects can be represented by a few keypoints since the background is almost static and the speakers do not change frequently. Therefore, instead of transmitting the whole video, one can share a source image with the receiver, and then only transmit the keypoints corresponding to each frame. While the source coding aspects of this problem are well-investigated, this article focuses on the effect of the transmission errors on keypoints. Transmission errors lead to expression changes in the reconstructed video, as opposed to complete loss of pixels in the conventional methods. The authors propose an incremental redundancy hybrid automatic repeat request framework (SVC-HARQ) that incorporates a novel semantic error detector. The authors also explore the use of channel state information for efficient keypoint transmission, which is shown to enhance the performance.

A semantic communication system for the transmission of speech signals over a wireless channel is studied by Han et al. in [A15], which uses text-related representation as the objective. An attention-based soft alignment module is used to extract text-related semantic features, while a speech reconstruction module is employed to recover the speech from text with limited additional information about the duration, pitch, power, and speaker identification. A final semantic correction module is used to correct the received text with semantic knowledge from the pretrained language model. The effectiveness of the proposed solution in both the text recognition performance and speech recovery is verified through numerical simulations.

The next article is a good example of the effectiveness of deep joint source-channel coding approaches beyond the transmission of classical information sources. In [A16], Xu et al. propose a deep joint source-channel coding framework for the channel state information (CSI) feedback task. Particularly, a non-linear transform method is developed to compress CSI and a signal-to-noise ratio (SNR) adaption mechanism is proposed to be adaptive to wireless channel variations. Extensive simulation results have been provided to verify the developed scheme.

Computation over multi-user networks can be considered as another type of end goal that goes beyond point-to-point goal-oriented communications. In [A17], Lin et al. study the problem of aggregation of local state information at devices that have to update their states. This is a key operation in distributed optimization, which concerns the optimization of a common function in a distributed network, and usually suffers from a communication bottleneck due the high number of messages that need to be exchanged. To mitigate this issue, the authors propose distributed over-the-air computing by exploiting simultaneous multicast beamforming and analog waveform superposition of a multiaccess channel. They consider two design criteria. The first one is to minimize the sum mean-squared error (MSE) with respect to the desired average-functional values. An efficient solution is proposed by transforming a non-convex beamforming problem into an equivalent concave-convex fractional program and solving it by nesting convex programming into a bisection search. The second criterion, called zero-forcing (ZF) multicast beamforming, is to force the received over-the-air aggregated signals at devices to be equal to the desired functional values. In this case, the optimal beamforming admits closed form. They analyze the convergence and show that the ZF beamforming outperforms the MSE design due to the bias in subgradient estimation of the latter. Simulations showcase that the proposed solutions also benefit from accelerated convergence by dramatically reducing the communication latency.

III. Pragmatic/Effective Communications

While all the above articles focus mainly on the reconstruction of an underlying source signal with some semantic-oriented reconstruction measure, in [A18], Meng et al. consider trajectory optimization over a time horizon. The synchronization between devices in the physical world and their digital models in the metaverse is crucial. This article proposes a sampling, communication and prediction co-design framework to synchronize the trajectories of a device in the real world and its digital model in the metaverse with minimal communication load. To optimize the sampling rate and the prediction horizon, the authors exploit expert knowledge and develop a constrained deep reinforcement learning algorithm. The framework is validated on a prototype composed of a real-world robotic arm and its digital model. The authors also show that with the assistance of expert knowledge, the proposed algorithm reduces the convergence time, the communication load, and the average tacking error.

Acknowledgment

The Guest Editors would like to express their gratitude to all the authors for their submissions, and all the anonymous reviewers for their insightful comments and questions that have contributed to the high quality of this Special Issue. The Guest Editors also acknowledge the strong support provided by Raouf Boutaba, Petar Popovski, Larry Milstein, and Janine Bruttin, that contributed to the success of this Special Issue.

Appendix: Related Articles

[A1] D. Gündüz et al., "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSA.2022.3223408.

[A2] H. Zou, C. Zhang, S. Lasaulce, L. Saludjian, and V. Poor, "Goal-oriented quantization: Analysis, design, and application to resource allocation," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221976.

[A3] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, "Toward semantic communications: Deep learning-based image semantic coding," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221999.

[A4] Z. Lei, P. Duan, X. Hong, J. F. C. Mota, J. Shi, and C.-X. Wang, "Progressive deep image compression for hybrid contexts of image classification and reconstruction," *IEEE J. Sel. Areas Commun.*, to be published, doi: 10.1109/JSAC.2022.3221998.

[A5] H. Wang et al., "VaBUS: Edge-cloud real-time video analytics via background understanding and subtraction," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221995.

[A6] P. Tandon et al., "Txt2Vid: Ultra-low bitrate compression of talking-head videos via text," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221953.

[A7] M. Kalfa et al., "Reliable extraction of semantic information and rate of innovation estimation for graph signals," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221950.

[A8] Y. Kim, J. Shin, Y. Cassuto, and L. R. Varshney, "Distributed boosting classifiers over noisy communication channels," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221972.

[A9] X. Mu, Y. Liu, L. Guo, and N. Al-Dhahir, "Heterogeneous semantic and bit communications: A semi-NOMA scheme," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3222000.

[A10] H. Zhang, S. Shao, M. Tao, X. Bi, and K. B. Letaief, "Deep learning-enabled semantic communication systems with task-unaware transmitter and dynamic data," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221991.

[A11] J. Kang et al., "Personalized saliency in task-oriented semantic communications: Image transmission and performance analysis," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221990.

[A12] C. Dong, H. Liang, X. Xu, S. Han, B. Wang, and P. Zhang, "Semantic communication system based on semantic slice models propagation,"'*IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221948.

[A13] S. Wang et al., "Wireless deep video semantic transmission," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221977.

[A14] P. Jiang, C. K. Wen, S. Jin, and G. Y. Li, "Wireless semantic communications for video conferencing," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221968.

[A15] T. Han, Q. Yang, Z. Shi, S. He, and Z. Zhang, "Semantic-preserved communication system for highly efficient speech transmission," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221952.

[A16] J. Xu, B. Ai, N. Wang, and W. Chen, "Deep joint source-channel coding for CSI feedback: An end-to-end approach," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221963.

[A17] Z. Lin, Y. Gong, and K. Huang, "Distributed over-the-air computing for fast distributed optimization: Beamforming design and convergence analysis," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3223661.

[A18] Z. Meng, C. She, G. Zhao, and D. De Martini, "Sampling, communication, and prediction co-design for synchronizing the real-world device and digital model in metaverse," *IEEE J. Sel. Areas Commun.*, 2022, doi: 10.1109/JSAC.2022.3221993.