WIRELESS SEMANTIC COMMUNICATIONS

CONCEPTS, PRINCIPLES, AND CHALLENGES

EDITED BY

YAO SUN | LAN ZHANG

DUSIT NIYATO | MUHAMMAD ALI IMRAN

WILEY

Wireless Semantic Communications

Wireless Semantic Communications

Concepts, Principles, and Challenges

Edited by

Yao Sun
University of Glasgow
Glasgow, UK

Lan Zhang
Michigan Technological University
Houghton, USA

Dusit Niyato
Nanyang Technological University
Singapore

Muhammad Ali Imran
University of Glasgow
Glasgow, UK

WILEY

This edition first published 2025

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data applied for:

Hardback: 9781394223305

Cover Design: Wiley
Cover Image: © Yuichiro Chino/Getty Images

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

Contents

List of Contributions

Xuyang Chen
College of Electronics and Information Engineering
Shenzhen University
Shenzhen
China

Runze Cheng
James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Xiangyi Deng
Glasgow College
University of Electronic Science and Technology of China
Chengdu
Sichuan
China

Hongyang Du
School of Computer Science and Engineering
Nanyang Technological University
Singapore

Daquan Feng
College of Electronics and Information Engineering
Shenzhen University
Shenzhen
China

Lei Feng
The State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing
China

Biqian Feng
Department of Electronic Engineering
Shanghai Jiao Tong University
China

Chenyuan Feng
Eurecom
France

Zhipeng Gao
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing
China

Shuaishuai Guo
School of Control Science and Engineering
Shandong University
Jinan
China

and

Shandong Key Laboratory of Wireless
Communication Technologies
Shandong University
Jinan
China

Qi He
National Key Laboratory of Science and
Technology on Communications
University of Electronic Science and
Technology of China
Chengdu
China

Chongwen Huang
College of Information Science and Electronic
Engineering
Zhejiang University
Hangzhou
Xihu Region
China

Muhammad Ali Imran
James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Rongpeng Li
College of Information Science and Electronic
Engineering
Zhejiang University
Hangzhou
China

Wenjing Li
The State Key Laboratory of Networking and
Switching Technology
Beijing University of Posts and
Telecommunications
Beijing
China

Chengsi Liang
James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Yijing Lin
State Key Laboratory of Networking and
Switching Technology
Beijing University of Posts and
Telecommunications
Beijing
China

Yijie Mao
School of Information Science and Technology
ShanghaiTech University
Shanghai
China

Fei Ni
College of Information Science and Electronic
Engineering
Zhejiang University
Hangzhou
China

Dusit Niyato
School of Computer Science and Engineering
Nanyang Technological University
Singapore

Yao Sun
James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Bingyan Wang
College of Information Science and Electronic Engineering
Zhejiang University
Hangzhou
China

Bohao Wang
College of Information Science and Electronic Engineering
Zhejiang University
Hangzhou
Xihu Region
China

Jiacheng Wang
School of Computer Science and Engineering
Nanyang Technological University
Singapore

Yanhu Wang
School of Control Science and Engineering
Shandong University
Jinan
China

and

Shandong Key Laboratory of Wireless Communication Technologies
Shandong University
Jinan
China

Yiwen Wang
School of Information Science and Technology
ShanghaiTech University
Shanghai
China

Le Xia
James Watt School of Engineering
University of Glasgow
Glasgow
Lanarkshire
UK

Xiang-Gen Xia
Department of Electrical and Computer Engineering
University of Delaware
Newark, DE
USA

Ruopeng Xu
College of Information Science and Electronic Engineering
Zhejiang University
Hangzhou
Xihu Region
China

Zhaohui Yang
College of Information Science & Electronic Engineering
Zhejiang University
Hangzhou
Xihu Region
China

Honggang Zhang
College of Information Science and Electronic Engineering
Zhejiang University
Hangzhou
China

and

Zhejiang Lab
Hangzhou
China

Xuefei Zhang
School of Information and Communication Engineering
Beijing University of Posts and Telecommunications
Beijing
China

Zhifeng Zhao
College of Information Science and Electronic Engineering
Zhejiang University
Hangzhou
China

and

Zhejiang Lab
Hangzhou
China

Jiakang Zheng
School of Electronic and Information Engineering
Beijing Jiaotong University
Beijing
China

Yu Zhou
The State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing
China

Preface

Tremendous traffic demands have increased in current wireless networks to accommodate for the upcoming pervasive network intelligence with a variety of advanced smart applications. In response to the ever-increasing data rates along with stringent requirements for low latency and high reliability, it is foreseeable that available communication resources like spectrum or energy will gradually become scarce in the upcoming years. Combined with the almost insurmountable Shannon limit, these destined bottlenecks are, therefore, motivating us to hunt for bold changes in the new design of future networks, i.e., making a paradigm shift from bit-based traditional communication to context-based semantic communication.

The concept of semantic communication was first introduced by Weaver in his landmark paper, which explicitly categorizes communication problems into three levels, including the technical problem at the bit level, the semantic problem at the semantic level, and the effectiveness problem at the information exchange level. Nowadays, the technical problem has been thoroughly investigated in the light of classical Shannon information theory, while the evolution toward semantic communication is just beginning to take shape, with the core focus on meaning delivery rather than traditional bit transmission.

Concretely, semantic communication first refines semantic features and filters out irrelevant content by encoding the semantic information (i.e., semantic encoding) at the source, which can greatly reduce the number of required bits while preserving the original meaning. Then, the powerful semantic decoders are deployed at the destination to accurately recover the source meaning from received bits (i.e., semantic decoding), even if there are intolerable bit errors at the syntactic level. Most importantly, through further leveraging matched background knowledge with respect to the observable messages between source and destination, users can acquire efficient exchanges for the desired information with ultralow semantic ambiguity by transmitting fewer bits.

While semantic communication offers these attractive and valuable benefits, it also faces many challenges. For example, when considering the computing limitation of terminal devices, personalized background knowledge, as well as unstable wireless channel conditions, how to design semantic encoder and decoder should be a challenging issue. Moreover, in the networking layer, it is nontrivial to seek the optimal wireless resource management strategy to optimize its overall network performance in a semantics-aware manner. Therefore, before fully enjoying the superiorities of semantic communications, this book would like to explore the following fundamental issues: (i) How many benefits can be achieved by using semantic communication? (ii) How much cost (mainly consumed resources) is incurred to guarantee the required performance, such as semantic ambiguity? and (iii) How can we maximize the benefits of semantic communications applied to different wireless networks with constraints of cost?

This book will explore recent advances in the theory and practice of semantic communication. In detail, the book covers the following aspects:

(1) Principles and fundamentals of semantic communication.
(2) Transceiver design of semantic communications.
(3) Resource management in semantic communication networks.
(4) Semantic communication applications to vertical industries and some typical communication scenarios.

Chapter 1 delves into the transceiver design for semantic communications. Specifically, we first summarize established designs for key components in semantic communications, with a key focus on the knowledge base and semantic encoders/decoders crucial for single-user semantic communications. Our discussion then extends to multiuser SC, specifically emphasizing the synergy between various multiple-access schemes and semantic communications, including orthogonal multiple access (OMA), space-division multiple access (SDMA), non-orthogonal multiple access (NOMA), rate-splitting multiple access (RSMA), and model-division multiple access (MDMA). In the end, we explore various applications of semantic communications and analyze the potential alterations in transceiver design required for these applications.

Chapter 2 studies semantic communications from a networking perspective, particularly focusing on the upper layer. Our primary objective is to investigate optimal wireless resource management strategies within the semantic communications-enabled network (SC-Net) to enhance overall network performance in a semantics-aware manner. This entails addressing the unique challenge of ensuring background knowledge alignment between multiple mobile users (MUs) and multitier base stations (BSs). Efficient resource management remains paramount within the SC-Net, offering numerous benefits such as guaranteeing high-quality SemCom services and enhancing spectrum utilization. By devising effective resource allocation strategies, we aim to optimize network performance and facilitate seamless communication within the SC-Net ecosystem.

Chapter 3 innovatively proposes the transformation theory of the semantic domain–spatial domain, projecting the knowledge graph onto a three-dimensional tensor in the spatial domain. By mapping the entity's semantic ambiguity with the intensity of discrete points, the knowledge graph is reconstructed from background knowledge libraries and three-dimensional tensors at the receiving end. Additionally, this chapter proposes a graph-to-graph semantic similarity (GGSS) metric based on graph optimal transport theory to evaluate the similarity of semantic information before and after transmission, as well as a semantic-level image-to-image semantic similarity (IISS) metric that aligns with human perception. Finally, we demonstrate the effectiveness and rationality of the framework through simulations.

Chapter 4 first introduces the problem that neural networks in semantic communication are very vulnerable to adversarial attacks, then proposes robust semantic communication systems for image and speech transmission. Meanwhile, this chapter discusses the privacy issue caused by the difference of knowledge bases between transmitter and receiver in semantic communication and proposes a knowledge discrepancy-oriented privacy protection (KDPP) method for semantic communication to reduce the risk of privacy leakage while retaining high data utility.

Chapter 5 investigates the means of knowledge learning in semantic communication with a particular focus on the utilization of Knowledge Graphs (KGs). Specifically, we first review existing efforts that combine semantic communication with knowledge learning. Subsequently, we introduce a KG-enhanced semantic communication system, wherein the receiver is carefully calibrated to leverage knowledge from its static knowledge base for ameliorating the decoding performance. Contingent upon this framework, we further explore potential approaches that can empower the system to operate in evolving an knowledge base more effectively. Furthermore, we investigate the possibility of integration with large language models (LLMs) for data augmentation, offering additional perspective into the potential implementation means of semantic communication. Extensive numerical results demonstrate that the proposed framework yields superior performance on top of the KG-enhanced decoding and manifests its versatility under different scenarios.

Chapter 6 presents a novel framework called VISTA (VIdeo transmission over Semantic communicaTion Approach) for video transmission by exploiting semantic communications. VISTA comprises three key modules: the semantic segmentation module and the frame interpolation module, responsible for semantic encoding and decoding, respectively, and the joint source-channel coding (JSCC) module, designed for SNR-adaptive wireless transmission.

Chapter 7 proposes a content-aware robust semantic communication framework for image transmission based on generative adversarial networks (GANs). Specifically, the accurate semantics of the image are extracted by the semantic encoder and divided into two parts for different downstream tasks: regions of interest (ROI) and regions of non-interest (RONI). By reducing the quantization accuracy of RONI, the amount of transmitted data volume is reduced significantly. During the transmission process of semantics, a signal-to-noise ratio (SNR) is randomly initialized, enabling the model to learn the average noise distribution. The experimental results demonstrate that by reducing the quantization level of RONI, transmitted data volume is reduced up to 60.53% compared to using globally consistent quantization while maintaining comparable performance to existing methods in downstream semantic segmentation tasks. Moreover, our model exhibits increased robustness with variable SNRs.

Chapter 8 first proposes an integrated framework for bridging meanings of semantic information in the Metaverse to achieve efficient interaction between physical and virtual worlds. This chapter then presents a Zero Knowledge Proof-based verification mechanism to secure the authenticity of the extracted information. This chapter also introduces a diffusion model-based resource allocation mechanism to maximize the utility of resources. Simulation results are presented to validate the authenticity and efficiency of the proposed mechanisms. Additionally, this chapter discusses future directions to further advance SemCom in the metaverse.

Chapter 9 discusses the method of leveraging large language model (LLM) to assist semantic communication systems. Specifically, this chapter first discusses leveraging LLM to define the semantic loss of communications, based on which a signal-shaping method is proposed to minimize the semantic loss for semantic communications with a few message candidates. Then, this proposed a more generalized method to quantify the semantic importance of a word/frame using LLM and investigate semantic importance-aware communications (SIAC) to reliably convey the semantics with limited communication and network resources. Finally, this chapter points out the future direction of using LLM for semantic correction. Experiments are conducted to verify the effectiveness of leveraging LLM to assist semantic communications.

Chapter 10 explores cutting-edge advancements in RIS for semantic communication. It delves into three pivotal areas: optimizing beamforming in RIS-aided systems for enhanced communication in complex digital environments like the Metaverse; employing physical layer strategies for robust privacy protection in semantic communication systems; and leveraging deep learning for advanced interpreting and prioritizing data and encoding and decoding semantic transmission in wireless communications. These topics collectively highlight the potential of RIS in transforming communication paradigms, emphasizing efficiency, security, and intelligent data processing in semantic communication.

Yao Sun
Lan Zhang
Dusit Niyato
Muhammad Ali Imran
April 2024