王江舟，中国工程院外籍院士、英国皇家工程院院士、IEEE Fellow、IET Fellow，英国肯特大学教授，研究领域是移动通信。他共发表了500多篇论文和5本专著，是2022年IEEE Leonard G. Abraham奖和IEEE GLOBECOM2012最佳论文奖的获得者，曾任IEEE Distinguished Lecturer、2019年IEEE国际通信大会（ICC2019）技术委员会主席、IEEE ICC2015执行主席、IEEE WCNC2013技术委员会主席，现任多个国际学术期刊编委，包括IEEE Transactions on Communications。

“语义通信的数学理论”序言

王江舟

客观世界由物质、能量和信息组成。探索和利用信息是信息时代的主要驱动力。回顾信息与通信的历史，如180年的经典通信和半个世纪的移动通信，我们看到了诸多通信技术的杰出成就，如超高速光通信、卫星通信、互联网、3G/4G/5G移动通信等。1948年，美国科学家香农发表了一篇著名论文，建立了经典信息论（CIT）[1]，以指导信息和通信系统的设计和优化。在过去的70多年里，人们开发了许多编码技术来接近香农指出的理论极限。例如，哈夫曼编码、算术编码和通用编码等，可以逼近数据压缩的极限，即信源熵。同样，近30年来，一些先进的信道编码，如Turbo码、LDPC码、Polar码等，已经接近或达到了信道容量，被视为信道编码理论的一大里程碑。此外，作为代表性方法，矢量量化、线性预测编码和变换编码可以趋近有损信源编码的率失真函数。总之，今天的信息与通信技术已经完全达到了经典信息论所预言的理论极限。

事实上，在信息和通信领域，我们生活在最好的时代和最坏的时代。对于普通人来说，无线互联网和4G/5G网络等通信技术使生活更轻松，甚至改变了社会形态。然而，对于专业人士来说，我们知道，由于经典信息论的停滞不前，通信技术的发展已经遇到了瓶颈。甚至一些悲观主义者认为，盛宴已经结束或物理层已经死亡，因为通信技术已经接近经典信息论的理论极限。那么，通信的未来方向是什么？回顾韦弗[2]的评论，他将信号处理中的信息分为三类，如语法信息（数据或消息的比特序列）、语义信息（消息的内容或意义）和语用信息（消息的价值或目的），并进一步指出经典信息论只处理语法信息。尽管以经典信息论为指导的语法通信系统已经逼近了理论极限，但语义信息处理是一个尚未开垦的领域，它提供了巨大的性能提升潜力。另一方面，在过去的10年中，人工智能（AI）中的机器学习和深度学习迅速发展，促进了自然语言处理（NLP）和计算机视觉（CV）中的许多应用。本质上，深度学习和神经网络模型的技术优势源于语义信息的利用。

通信与人工智能的关系如图1所示，物理世界的人机物等一切对象的关系与交互都会被映射到数字世界。在数字世界中，通信系统采集与传输语法信息，在算力、数据以及算法支持下，人工智能技术从语法信息中提取语义信息进行加工处理，进一步，智能决策与控制系统基于语用信息对物理世界执行动作。由此可见，语义信息是通信与人工智能融合的关键，借助语义信息，通信融合AI将实现“最后一公里”的目标，推动无线AI、具身智能等应用的成熟与普及。信息技术的发展迫切需要一个语义信息的成熟理论，它将极大地促进通信和人工智能技术的快速进步。

我向广大读者热切推荐牛凯教授与张平院士撰写的“语义通信的数学理论”这篇文章。在该文中，作者构建了一个语义信息理论（SIT）的数学框架。作为一个合理可行的理论体系，语义信息论

基金项目：国家自然科学基金资助项目（No.62293481, No.62071058）

应该是经典信息论的自然延伸，后者可以作为特例纳入前者。此外，语义信息论应当回答3个基本问题，即什么是语义信息，如何测量语义信息，以及为什么语义通信系统具有性能增益。在这篇论文提出的语义信息论框架中，所有这些问题都得到了明确的结论。

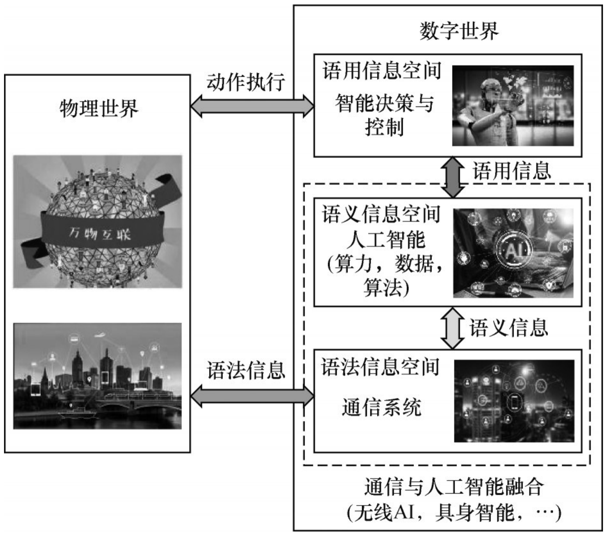

图1　通信与人工智能的关系

首先，这篇论文研究了语义信息的本质。从概念上讲，语义信息是指隐藏在数据或消息背后的含义或内容。历史上，许多工作探讨了语义信息的性质，并给出了各种定义，如基于逻辑概率的语义信息概念和由模糊集定义的概念等。然而，前人的大多数定义和概念只揭示了语义信息的部分属性，并且只适用于一些有限的场景。该文指出语义信息是语法信息的上级概念，是许多等效或相似语法信息的抽象表征。因此，在论文中使用同义性（语言学的一个术语）来表示语义信息的这种抽象特征。此外，作者将语义信息与语法信息之间的关系命名为同义映射，这是一种“一对多”映射，即一个语义符号可以由许多不同的语法符号表示。论文指出，在以往的许多工作中，语义信息的表示，如逻辑概率、知识表示、特征图、下游任务等，都可以用同义映射来表示。因此，同义映射是语义信息的本质，它揭示了语法信息的特征提取和意义抽象。

其次，基于同义映射，作者引入语义信息的完整度量体系，包括语义熵$H_s(\tilde{U})$、上/下语义互信息$I^s(\tilde{X};\tilde{Y})(I_s(\tilde{X};\tilde{Y}))$、语义信道容量$C_s$以及语义率失真函数$R_s(D)$。在所有度量中，语义熵$H_s(\tilde{U})$是核心指标，表示为信源概率分布与同义映射的泛函。语义熵$H_s(\tilde{U})$不同于信息熵$H(U)$，它由同义映射约束，如果采用“一对一”映射，则可以退化为传统的信息熵。因此，作者得出结论，语义熵是信息熵的自然延伸。进一步，该文还定义了2个随机变量和整个随机序列的同义映射，以引入条件熵和联合语义熵。相应地，论文讨论了序列语义熵的链式法则。特别地，作者引入了上/下语义互信息，这是语义信息论和经典信息论之间的主要区别。在经典信息论中，利用互信息的凹凸特性来评估信道容量和率失真函数。相反，在语义信息论中，需要对信源分布和联合同义映射优化，得到上语义互信息的最大值，从而求得语义信道容量；另一方面，需要对测试信道转移概率分布和信源/信宿的同义映射进行优化，得到语义互信息的最小值，从而求得语义率失真函数。此外，论文证明了语义熵（语义率失真函数）不大于信息熵（经典率失真函数），即$H_s(\tilde{U})\leqslant H(U)$与$R_s(D)\leqslant R(D)$，以及语义信道容量不小于经典信道容量，即$C_s\geqslant C$。这些结论中，当采用“一对一”映射时，所有不等式的等号成立。由此可见，语义信息度量体系可以包含语法信息度量，前者与后者具有兼容性。

第三，论文证明了3个重要的语义编码定理，以揭示语义通信的性能优势。在经典信息论中，香农的主要贡献是通过使用渐近均分性质（AEP）和（联合）典型序列译码/编码来证明经典编码定理。无失真/限失真信源编码定理和信道编码定理在经典通信中起着核心作用，指出了语法通信的性能极限。类似地，基于同义映射，作者引入了新的数学工具，即语义AEP和同义典型序列译码/编码，以证明语义编码定理，如语义无失真信源编码定理、语义信道编码定理和语义限失真信源编码定理。类似于经典信息论，这些基本编码定理也都是存在性定理，如何构造最优语义编码方案仍然是开放问题，但它们约束了语义通信系统的性能边界，并在语义信息论中起着关键作用。由此得出结论，语义

容量大于经典容量，语义压缩率（语义熵和语义率失真）小于经典度量。从理论上讲，由于同义映射和基本编码定理，语义通信系统可以优于经典通信系统。

最后，论文讨论了连续情况下的语义信息度量。在这里，同义映射被转换为连续随机变量分布区间的划分方式。相应地，划分后的子区间被命名为同义区间，其平均长度被定义为同义长度S。此外，论文推导了连续条件下的语义信息度量体系，包括语义熵、上/下语义互信息、语义信道容量和语义率失真函数。论文发现区间划分方法和同义区间长度将影响所有这些语义信息度量。特别是对于限带高斯信道，作者得到了一个新的信道容量公式$C_s = B\log\left[S^4\left(1+\frac{P}{N_0 B}\right)\right]$，其中同义长度$S$表示了算力与算法的处理能力。当$S = 1$时，该公式退化为著名的香农信道容量公式。由此可见，这一容量公式是经典信道容量的重要扩展，揭示了语义通信性能提升的巨大潜力。

在物理学史上，基于光速不变的假设，爱因斯坦建立了狭义相对论，将经典力学作为特例包含进来。类似地，这篇论文建立了一个源自同义映射概念的语义信息数学框架。由于语义信息论与经典信息论兼容，因此它是一个完整自洽的理论。应用该理论，可以系统地测量和评估语义信息，并设计和优化语义通信系统。论文还证明了3个语义编码定理，并探索了语义通信的基本极限，即语义信道容量和语义熵/率失真。

一个伟大的理论必然能概括为系统化的数学表述，而这种数学表述越简洁，就越容易被接受，必将引导一场技术革命，从而形成一种创新的社会范式。在历史上，经典信息论作为一种开创性的理论，揭开了信息技术革命的面纱。特别是香农的信道容量公式，成为通信系统设计与优化的指路明灯。展望未来，语义信息论将指导语义通信系统的优化，设计最优的语义编码方案，将为未来的通信和信号处理开辟新的道路。

参考文献：

[1] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27: 379-423, 623-656.

[2] WEAVER W. Recent contributions to the mathematical theory of communication[J]. ETC: A Review of General Semantics, 1953, 10(4): 261-281.

Prolog of a mathematical theory of semantic communication

Jiangzhou Wang

The objective world consists of matter, energy, and information. Exploring and utilizing information is the major driving force of the information era. Retrospecting the history of information and communication, e.g., 180 years of classic communications and half a century of mobile communications, we observe the outstanding achievements of communication technologies, such as ultra-high-speed optical communication, satellite communication, internet, 3G/4G/5G mobile communication and so on. In 1948, C. E. Shannon published a famous paper and established the classic information theory (CIT)[1] to guide the design and optimization of information and communication systems. For the past 70 years, we have developed many coding techniques to approach the theoretic limits pointed out by Shannon. For instance, Huffman coding, arithmetic coding, and universal coding can approach the limitation of data compression, that is, the information entropy. Similarly, in the last thirty years, some advanced channel codes, such as turbo code, LDPC code, and polar code, etc., have approached or achieved the channel capacity, which is regarded as a great milestone of coding theory. Furthermore, as the representative methods, vector quantization, linear prediction coding, and transform coding can be close to the rate-distortion function of lossy source coding. In a word, the information and communication techniques today have roundly reached the theoretical limitation predicted by the classic information theory.

In fact, in the field of information and communication, we live in the best of times and the worst of times. For ordinary people, communication technology, such as wireless internet and 4G/5G network has made life easier and even changed the form of society. However, for professionals, we are aware that the development of communication technology has reached a bottleneck due to the stagnation of classic information theory. Some pessimists believe that the party is over or the physical layer is dead since the communication techniques have approached the theoretic limits of CIT. So, what is the future direction of communication? Looking back the comment of Weaver in Ref. [2], he divided the information in signal processing into three categories, such as syntactic information (the bit sequence of data or message), semantic information (the content or meaning of a message), and pragmatic information (the value or purpose of a message), and further pointed out that the CIT only handles the syntactic information. Although the syntactic communication systems guided by the CIT have reached the theoretic limitation, semantic information processing is uncultivated territory that offers tremendous potential for performance gains. In the last decade, machine learning and deep learning in artificial intelligence (AI) have quickly developed to promote many applications in natural language processing (NLP) and computer vision (CV). Essentially, the technical advantages of deep learning and neural network models of NLP or CV stem from the utilization of semantic information.

The relationship between communication and artificial intelligence is shown in Fig. 1. In the physical world, the relationships and interactions among all entities, such as humans, machines, and objects, are mapped to the digital world, in which the communication system collects and transmits syntactic information, and with the support of computing power, data and algorithms, artificial intelligence technology extracts semantic information from syntactic information for processing, and furthermore, the intelligent decision and control system performs actions on the physical world based on pragmatic information. It can be seen that semantic information is the hinge to merge communication and artificial intelligence, and with the help of semantic information, communication integrated AI will achieve the goal of “last mile” and promote the maturity and popularization of applications such as wireless AI and embodied intelligence, etc. The development of information technology urgently needs a mature theory of semantic information, which will greatly promote the rapid progress of communication and artificial intelligence technology.

I eagerly recommend the paper “A Mathematical Theory of Semantic Communication” written by Prof. Kai Niu and Prof. Ping Zhang to the reader. In this paper, the authors try to establish a mathematical framework of semantic information theory (SIT). As a feasible framework, SIT should be a natural extension of CIT and the latter can be included in the former as a special case. Furthermore, SIT should answer three primary problems, that is, what is the semantic information, how to measure the semantic information, and why the semantic communication systems have performance gains. In their framework of SIT, all these problems are provided with a definite conclusion.

This work is supported by the National Natural Science Foundation of China (No.62293481, No.62071058).

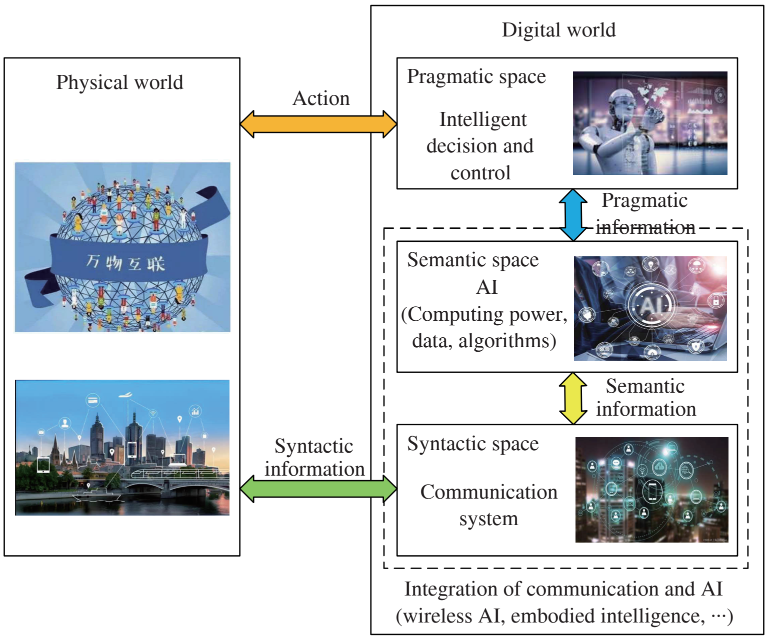

Fig. 1 The relationship between communication and AI

First, this paper investigates the essence of semantic information. Conceptually, semantic information refers to the meaning or content that is hidden behind data or messages. In history, many works explored the property of semantic information and gave various definitions, such as the semantic information concept based on logical probability and the concept defined from the fuzzy set, etc. However, most of the existing definitions and concepts reveal partial properties of semantic information and are only suitable for some limited scenarios. The authors realize that semantic information is a high-level concept of syntactic information and an abstract presentation of many equivalent or similar syntactic messages. Therefore, in this paper, they use the synonym, a terminology of linguistics, to indicate such an abstract presentation of semantic information. Furthermore, they name the relationship between the semantic information and the syntactic information as the synonymous mapping, which is a "one-to-many" mapping, that is, one semantic symbol can be presented by many different syntactic symbols. They find that many presentations of semantic information in the existing works, such as logical probability, knowledge presentation, feature map, downstream task, etc. can be indicated by the synonymous mapping. Therefore, synonymous mapping is the essence of semantic information which reveals the feature extraction and meaning abstraction of syntactic messages.

Second, based on the synonymous mapping, the authors introduce a full measurement system of semantic information including semantic entropy $H_s(\tilde{U})$, up and down semantic mutual information $I^s(\tilde{X},\tilde{Y})$ $(I_s(\tilde{X},\tilde{Y}))$, semantic channel capacity C_s, and semantic rate-distortion function $R_s(D)$. In all measures, the semantic entropy $H_s(\tilde{U})$ is the core target, which is a functional of the source probability distribution and the synonymous mapping. The semantic entropy $H_s(\tilde{U})$ is different from the information entropy $H(U)$ since it is also constrained by the synonymous mapping but this target can be degraded into the traditional information entropy if a "one-to-one" mapping is applied. Hence, they conclude that semantic entropy is a natural extension of the information entropy. They also define the synonymous mapping over two random variables and a random sequence to introduce the conditional and joint semantic entropy. Correspondingly, the authors discuss the chain rule of sequential semantic entropy. In particular, they introduce the up/down semantic mutual information, which is the main difference between SIT and CIT. In the CIT, the concave and convex properties of mutual information are used to evaluate the channel capacity and rate distortion. On the contrary, in the SIT, the maximum of up semantic mutual information over the source distribution and the jointly synonymous mapping indicates the semantic channel capacity and the minimum of down semantic mutual information over the transition probability distribution of the test channel and the source/destination synonymous mapping presents the semantic rate-distortion. In addition, the authors prove that the semantic entropy (semantic rate-distortion) is no more than the information entropy (classic rate-distortion), that is, $H_s(\tilde{U}) \leqslant H(U)$ and $R_s(D) \leqslant R(D)$, and the semantic channel capacity is no less than the classic counterpart, that is, $C_s \geqslant C$, where all the equalities hold if a "one-to-one" mapping is used. It follows that the measurement system of semantic information can contain the measures of syntactic information as the special cases and the former is compatible with the latter.

Third, the authors prove three important semantic coding theorems to uncover the performance advantages of semantic communication. In the CIT, the main contribution of Shannon is to prove the classic coding theorems by using the asymptotic equipartition property (AEP) and (jointly) typical sequence decoding/encoding. The lossless/lossy source coding theorem and channel coding theorem play core roles in classic communication and provide the performance limitation of syntactic communication. Similarly, based on the synonymous mapping, the authors introduce new mathematical tools, that is, the semantic AEP and synonymous typical sequence decoding/encoding, to prove the semantic coding theorems, such as semantic lossless source coding theorem, semantic channel coding theorem, and semantic lossy source coding theorem. Similar to classic counterparts, these primary coding theorems are all existence theorems, and how to construct the optimal coding schemes is still an open problem. Nevertheless, they constrain the performance boundary of the semantic communication system and play critical roles in the SIT. Hence, this paper concludes that the semantic channel capacity is larger than the classic counterpart and the semantic compressive rates (semantic entropy and semantic rate-distortion) are less than the traditional measures. Theoretically, due to the synonymous mapping and the basic coding theorems, the semantic communication system can outperform the classic communication one.

Finally, the authors discuss the semantic information measures in the continuous case. Here, the synonymous mapping is transformed into a partition of continuous intervals. Correspondingly, the partition subinterval is named as the synonymous interval and its average length is defined as the synonymous length S. Furthermore, they derive the measurement system under continuous conditions, including semantic information entropy, up/down semantic mutual information, semantic channel capacity, and semantic rate-distortion. They find that the partition method and the associated synonymous length will affect all these semantic information measures. Especially, for the band-limited Gaussian channel, a new channel capacity formula is derived, that is, $C_s = B\log[S^4(1+\frac{P}{N_0B})]$, where the synonymous length S indicates the handling ability of computing power and algorithm processing. When $S=1$, this formula is transformed into the famous Shannon formula whereby it can be regarded as an important extension of classic capacity and reveals the great potential for performance improvement of semantic communication.

In the history of physics, based on the assumption that the speed of light does not change, Einstein created the special theory of relativity, which included classical mechanics as a special case. Similarly, this paper establishes a mathematical framework of semantic information stemming from the concept of synonymous mapping. Since semantic information theory is compatible with classical information theory, it is a complete and self-consistent theory. By using this theoretic framework, the semantic information can be systematically measured and evaluated and the semantic communication system can be designed and optimized. The authors also prove three semantic coding theorems and explore the fundamental limits of semantic communication, that is, semantic channel capacity and semantic entropy/rate distortion.

A great theory must be summarized as a systematic mathematical representation, and the more concise this mathematical representation is, the more acceptable it is, and it will inevitably guide a technique revolution, thus forming an innovative social paradigm. In history, classical information theory, as a pioneering theory, has unveiled the revolution of information technology. In particular, Shannon's channel capacity formula has become a guiding light for the design and optimization of communication systems. Looking forward to the future, semantic information theory will guide the optimization of semantic communication systems. Designing the optimal semantic coding will open up a new path for future communication and signal processing.

REFERENCES

[1] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27: 379-423, 623-656.
[2] WEAVER W. Recent contributions to the mathematical theory of communication[J]. ETC: A Review of General Semantics, 1953: 261-281.

ABOUT THE AUTHORS

Jiangzhou Wang (Fellow, IEEE) is a Professor with the University of Kent, U.K. He has published more than 500 papers and five books. His research interest is in mobile communications. He was a recipient of the 2022 IEEE Communications Society Leonard G. Abraham Prize. He was the Technical Program Chair of the 2019 IEEE International Conference on Communications (ICC2019), Shanghai, Executive Chair of the IEEE ICC2015, London, and Technical Program Chair of the IEEE WCNC2013. He is/was the editor of multiple international journals, including IEEE Transactions on Communications from 1998 to 2013. Professor Wang is an International Member of the Chinese Academy of Engineering (CAE), a Fellow of the Royal Academy of Engineering (RAEng), U.K., Fellow of the IEEE, and Fellow of the IET.