

面向万物智联的语义通信网络

石光明^{1,3}, 肖泳^{2,3}, 李莹玉², 高大化^{1,3}, 谢雪梅^{1,3}

(1. 西安电子科技大学人工智能学院, 陕西 西安 710071;
2. 华中科技大学电子信息与通信学院, 湖北 武汉 430074; 3. 琶洲实验室, 广东 广州 510330)

摘要: 随着无线通信智能化应用需求的快速提升, 未来通信网络将从单纯追求高传输速率的传统架构向面向万物智联的全新架构转变。语义通信是一种可将用户的需求和信息含义融入通信过程中的全新架构, 该架构有望成为未来万物智联网络的新型基础范式。介绍了语义通信与万物智联网络的关系及语义通信的基本模型和组成。通过分析点对点语义通信的局限性, 提出以知识共享和资源融合为基础的语义通信网络将更适合成为大规模智联网络的基础; 通过分析语义通信网络的基本组成, 介绍了一种基于联邦边缘智能的语义通信网络范例。仿真结果显示, 语义通信网络将有望在保证数据安全的同时, 大幅度降低通信的资源需求量并提高通信效率。最后, 探讨了语义通信网络未来发展的开放性问题。

关键词: 语义通信; 万物智联; 边缘智能; 联邦学习

中图分类号: TN929.5

文献标识码: A

doi: 10.11959/j.issn.2096-3750.2021.00209

Semantic communication networking for the intelligence of everything

SHI Guangming^{1,3}, XIAO Yong^{2,3}, LI Yingyu², GAO Dahua^{1,3}, XIE Xuemei^{1,3}

1. School of AI, Xidian University, Xi'an 710071, China
2. School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China
3. Pazhou Lab, Guangzhou 510330, China

Abstract: With the fast growing demand for the wireless network intelligence, the future communication system will transform from the traditional data-oriented solution to a novel intelligence-of-everything (IoE)-based architecture. Semantic communication is a new communication technology which involves the meaning of message into the communication process. It is believed that semantic communication will have the potential to serve as the fundamental paradigm for the future IoE. The relationship between the semantic communication and IoE was discussed and the basic models and fundamental components of semantic communication were introduced. By discussing the limitations of point-to-point semantic communication, it was argued that the knowledge sharing and resource convergence-based semantic communication networking would be ideal for supporting the future massive scales of IoE systems. The basic components of the semantic communication networking system were discussed and a federated edge intelligence-based semantic communication networking architecture as a case study was considered. Simulation results show that semantic communication networking has the potential to further reduce the resource demand and improve the efficiency of semantic communication. Finally, open problems for future research were discussed.

Key words: semantic communication, intelligence of everything, edge intelligence, federated learning

收稿日期: 2021-01-11; 修回日期: 2021-02-04

通信作者: 石光明, gmshi@xidian.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61621005, No.61632019, No.61836008, No.62071193); 中国博士后科学基金面上项目 (No.2020M672357)

Foundation Items: The National Natural Science Foundation of China (No.61621005, No.61632019, No.61836008, No.62071193), China Postdoctoral Science Foundation (No.2020M672357)

1 引言

过去 30 年，在人们对移动互联网和无线数据业务不断增长的需求的推动下，无线通信领域得到了迅速发展。3G 提供了可收发短信和进行基本网页浏览的无线数据服务。4G 大幅度提升传输速率，使移动视频流媒体业务得以普及。5G 则针对海量垂直业务的不同需求，首次引入全新的基于服务架构 (SBA, service-based architecture)，旨在实现通信网络从单纯追求高传输速率的传统架构向可支持三大业务场景和多个垂直行业的全新架构转变，为增强现实/虚拟现实技术、大规模物联网和自动驾驶等新兴应用场景的推广和普及提供保障。在未来 6G 时代，当所有物理层维度资源均接近饱和状态时，如何进一步提高通信效率，持续满足人们复杂、多样和智能化信息传输的需求，是无线技术发展面临的新挑战^[1]。

近年来，随着人们对无线通信智能化需求的迅速提升，各种基于无线通信技术的新兴智能业务（如工业互联网、智能网联车、远程医疗/手术、虚拟现实及全息投影技术等）层出不穷。这些新兴业务不再仅仅依靠高速率的数据传输，而逐渐对网络智能化和服务多样性等方面提出更多要求。在这一发展趋势的推动下，未来通信网络将开始逐步向高度自动化、智能化且可提供更贴近人类用户需求和体验的万物智联全新架构转变^[2]。

语义通信是一种全新的通信架构，它通过将用户对信息的需求及语义融入通信过程，将有望大幅度提高通信效率、改进用户的体验质量 (QoE, quality of experience)，并从根本上解决基于数据的传统通信协议中存在的跨系统、跨协议、跨网络、跨人一机不兼容和难互通等问题，真正实现“万物透明智联”的宏伟愿景，即通信网络、计算和存储等软/硬件设备无缝融入人们的生活。人们无须携带手机或计算机等专属计算和通信设备也可以享受无处不在的计算、存储和通信等服务^[3]。

本文从面向未来的万物智联网络出发，首先介绍语义通信与万物智联网络之间关系，并通过分析通信的内涵，阐明语义通信将有可能成为未来通信技术发展新范式的趋势。随后，介绍语义通信的基本模型和组成。通过分析点对点语义通信的局限性，本文认为以知识共享和资源融合为基础的语义通信网络更适合成为万物智联网络的基础。进一步

地，分析了语义通信网络的基本组成，并介绍了一种基于联邦边缘智能的语义通信网络组网范例。仿真结果显示，语义通信网络将有望在保证数据安全的同时，大幅度降低资源占用率并提高通信效率。最后，探讨了语义通信网络未来发展的开放性问题。

2 万物智联网络与语义通信网络

未来 6G 时代将延续 5G 向多个垂直行业延伸这一趋势，并面向更多更复杂的业务场景，为海量用户提供更多样化及个性化的服务^[2]。在该发展趋势的推动下，未来通信网络将从单纯追求高传输速率的传统架构向可支持海量设备和多种智能化业务场景，具备智能感知、智能学习、智能决策和智能演进的全新万物智联架构转变。具体而言，万物智联网络将包含以下 3 种类型的交互通信模式。

1) 机—机 (M2M, machine to machine) 智联主要针对海量物联网设备，如低功耗传感器及工业互联网中的仪器仪表设备等。目前，已有大量无线通信协议和接口（如蓝牙、Wi-Fi、LTE/4G 和 5G 等）支持不同类型物联网设备之间的数据连接。然而，不同网络、协议和传输技术之间一般无法实现数据互通。虽然近年来出现了一些跨协议数据互通技术，但目前仍缺乏简单且通用的跨网络、跨协议的数据互通技术和方法^[4]。随着各种新兴协议和通信技术（如 Google 的 QUIC、Wi-Fi 6 和 Wi-Fi 7 等）的出现，未来通信网络大量采用不同技术，且互不兼容协议共存的场景范围将进一步扩大。M2M 智联旨在实现不同技术与协议之间的自动切换，并将智能交互与智能决策有机结合，实现跨设备、跨资源和跨模态协同感知与合作决策。

2) 人—机 (H2M, human to machine) 智联主要包括可满足用户体验和个性化需求的人—机交互通信技术和接口。H2M 智联一般针对多种传感器和虚拟或实体人—机交互设备，并高度依赖基于人工智能技术的感知、学习和自适应算法对人类用户的情绪、个性等进行识别与感知。随着触觉互联网、人—机交互网络等新兴应用的出现，未来通信网络中将会包含大量复杂的 H2M 智联应用和服务^[5]。

3) 人—人 (H2H, human to human) 智联包含不同用户之间通过通信网络实现接近真实情景的信息、情感、感知等方面的交互。随着人们对网络应用和服务需求的不断提高，传统的以文本、语音

和图像为主要载体的人一人交互已经不能满足需求,而 3D 立体全息投影、触觉互联网等大量新兴 H2H 智联技术将会逐渐普及并在未来通信网络中占据越来越重要的地位。

尽管上述智联接口与技术可以极大地改善用户体验,并大幅扩展通信网络业务范围和应用场景,但是,目前尚缺乏可统一上述 3 种智联方式的归一化通信网络架构。此外,随着大量新兴业务的出现、推广和普及,现有网络架构将难以满足不同智联方式对资源业务用户需求多样化等方面快速增长的需求。

作为一种全新通信范式,语义通信技术有望将通信网络从传统的基于数据协议和格式的单一固化通信架构中解放出来,通过采用更具有普适性的信息含义作为衡量信息通信性能的主要指标,打通 M2M 智联、H2M 智联与 H2H 智联模式之间的“壁垒”,实现真正的“透明智联”。具体地,由于语义通信主要依赖于建立在人类用户和机器之间都具备普适性和可理解性的语义知识库,因此,有望打破目前 M2M 智联中信息模态不一致导致的不兼容性问题,为建立能够满足不同类型设备之间互通互联的统一通信协议架构奠定基础。其次,由于语义通信以人类的普适性知识和语义体系作为基础,因此,可从根本保证 H2M 智联与 H2H 智联交互及通信时的用户体验,并进一步减少语义和物理信号之间的转换次数,从而降低可能产生的语义失真。

3 语义通信的研究现状

语义通信的基本概念由 Weaver 和 Shannon 在 1948 年发表的经典论文中首次提出^[6],随后,学术界和工业界均从不同角度开展广泛研究。其中,学术界主要以 Shannon 信息论为基础开展相关研究。具体而言,Shannon 信息论的核心是采用传输信号的不确定性定义信息量,即以最小信息单位为“比特”的量对发送的信息是“0”或“1”的不确定性概率进行衡量^[7-8],并采用“熵”量化每秒发送信息的不确定性概率。正如 Weaver 和 Shannon 在其论文中所描述的,Shannon 信息论的基本方法可以直接延伸至用于衡量语义的信息量,即发送的符号不再是二进制的“0”和“1”,而是可能发送的语义信息符号。这一结论引起了学术界对语义通信的关注。例如,1952 年,Carnap 和 Bar-Hillel 采用概率

逻辑理论对如何衡量句子中所包含的语义信息量进行了研究^[9],即每一个句子(所对应的语义)为“真”的逻辑概率越大,则所包含的信息量(即熵的值)越小。其中,作者将语义符号(用 A 或者 B 表示)表示一个可为“是”或者“非”的事件,该事件为“是”与“非”的概率决定了语义的熵值大小,用 $H(A)$ 表示。2004 年,Floridi 发现 Carnap 和 Bar-Hillel 建立的语义通信理论存在一个基本悖论,称为 Carnap-Bar-Hillel 悖论(Carnap-Bar-Hillel Paradox),即任何事实与其自身矛盾时都会拥有无穷大的信息量,如熵 $H(A \wedge \neg A)$ (表示“事件 A 与事件非 A”发生的概率)等于无穷大^[10]。为了解决这一问题,Floridi 提出语义信息量应当采用与真实事件之间的距离表示,由此激发了一系列采用语义距离度量信息的方法。尽管采用语义的相对距离可以有效地避免出现 Carnap-Bar-Hillel 悖论情况,但是这些方法均需要一个为“真”的语义事件作为参照物;当该参照事件不存在时,则无法准确定义语义的信息量。2011 年,Bao 采用命题逻辑理论将 Carnap 和 Bar-Hillel 的工作做了进一步的推广^[11]。

近年来,语义通信技术引起了我国学者的关注^[3,12-16]。北京邮电大学钟义信认为语义信息理论的重点在于对内容的理解和逻辑演绎能力^[14-15]。北京邮电大学的张平也指出,语义通信将在 6G 技术中扮演重要角色,并最终实现人与网络节点融为一体,共同演进^[16]。

语义通信技术也引起了工业界的广泛关注。1992 年,美国软件公司 Business Objects 注册并获得了建立通用语义层(Semantic layer)的专利申请。该专利建议在数据和用户之间建立一个通用的语义层,用于自动转换和识别数据中包含的已在客户和产品之间熟悉的商业术语,从而在不同的商业客户与终端用户之间建立公共的术语沟通与表达方式。此外,美国思科公司在 2017 年提出了意图驱动网络概念,通过自动识别不同的网络域(如不同的子网络、广域网、数据中心及混合云计算中心等)用户、设备、应用和服务之间的需求或意图,据此对网络架构及工作流程进行自动控制、管理和优化。

4 通信的基本问题

Shannon 在其发表的经典论文中将通信的基本问题定义为“将信息从一个点准确或者近似地复制

到另一个点”^[8]。Shannon 认为，虽然通信系统中传输的信息通常包含一定的语义信息，但是这些语义信息一般受到诸多与通信物理信道及传输技术本身无关因素的影响。因此，为保证信息理论的普适性，在设计和考虑通信的技术问题时，不应当考虑通信的语义问题。

受到 Shannon 经典论文启发，Weaver 和 Shannon 在 1948 年发表的论文中提出了普适通信 3 个层次的问题，依次描述如下^[6]。

- 1) 通信的技术问题：所传输的通信符号是否准确？
- 2) 通信的语义问题：所传输的符号是否准确地表达了预期含义？
- 3) 通信的有效性问题：所接收的信息含义能否按照预期方式影响用户行为？

由于 Shannon 经典信息论中去除了语义对通信的影响，因此，主要解决了通信的技术问题，即通过引入熵的概念解决发送信号所能包含的最大信息量问题，并通过引入信道容量的概念解决通信信道数据传输量的极限问题^[7]。在 Weaver 和 Shannon 的论文中还列举了如下例子用以说明技术问题和语义问题之间的区别：“通信的技术问题就好比一个严肃的女生接收一个电报，她忽略电报信息所包含的任何语气，如语气是否悲伤、愉悦或者尴尬等”。

与技术问题不同，通信的语义问题不仅与传输的信号及信道的物理特征有关，还与信号所包含的信息类别、收发双方对消息的理解与识别准确度及接收端用户收到信息的满意度等复杂因素有关。有效性问题则考虑了通信更深层次问题，即任何信息的通信与交流都可以理解成为尝试影响用户行为的手段。有效的通信可能对用户长期或短期行为产生较为明显的影响，而无效的通信则可能无法对用户行为造成任何影响或者仅造成有限影响。Weaver 和 Shannon 都注意到通信的语义和有效性问题都需要充分借助除通信领域以外的知识和研究成果，包括语言、写作和艺术等交流与表达方式本身的特征和形式，以及来自心理学、哲学和人类学等诸多领域的知识和广泛成果。例如，通信的参与者需要共享相同或者相似的哲学世界观与知识模型，其中就可能牵扯本体论、知识局限性、认识论、逻辑和美学等在内的多个跨学科研究领域。本文主要聚焦通信的语义问题。

5 点对点语义通信模型、组成及局限性

5.1 语义通信模型

目前，语义通信的主要模型可大致分为两类，即目标驱动的通用语义通信模型和基于 Shannon 信息论的语义通信模型。

5.1.1 目标驱动的语义通信模型

2010 年，Juba 在其博士论文中首次提出了“通用语义通信”的概念^[17]。该概念的核心思想是在信息发射端和接收端缺乏公共语言或通用协议的情况下，通过感知进行目标沟通，从而实现基于目标的通用语义通信^[18-19]。该工作主要解决了传统通信系统中协议不兼容导致的通信不畅问题。作者通过建立通用的目标驱动语义通信数学模型，研究在相互不兼容的协议之间进行有限和无限目标沟通的可行性，同时对协议的不兼容性与实现通用语义理解和目标沟通复杂度之间的关系进行了量化建模。该文首次证明了感知是实现通用语义通信的充分必要条件。最后，作者以 TCP 的变种为例，验证了包交换网络中，在不同包大小和格式协议之间实现有效通信的可行性。值得注意的是，尽管作者强调发射端与接收端可以在具备不同背景的情况下通过通用语义通信模型进行沟通，但是该模型仍然要求收/发两端具有对于通信是否成功的统一参考标准。因此，该架构所阐述的收/发两端可具备不同背景，主要是指语言和沟通协议不同，并非常识背景知识库或通识性语义模型不同。

5.1.2 基于 Shannon 信息论的语义通信模型

Weaver 和 Shannon 在 1948 年发表的经典论文中阐明，语义通信的基本问题可以通过对 Shannon 经典信息论基础框架的延伸得到^[6]。具体而言，Shannon 信息论采用不同信息码字的概率分布（即不确定性）表示信息量大小；信息不确定性越大，熵也就越大，所包含的信息量越大。可通过对类似方法进行拓展表达语义通信的信息量，即以可能出现的语义信息符号的概率分布作为语义信息量的度量——语义信息的不确定性越大，则语义信息的熵也就越大。可采用如下数学模型进行描述，定义 \mathcal{X} 为可能发送的有限码字集合，用 X 表征可能发送码字的一个离散随机变量（如 0 和 1 的数字序列）。相似的，定义 \mathcal{S} 为可能发送的有限语义信息集合，用 S 表征源信号中可能出现语义信息的离散随机变量（如发射端用户希望传输的一个逻辑事件）。

采用 Shannon 信息熵 $H(X)$ 和 $H(S)$ 衡量发送码字和语义信息的信息量, 即发送码字和语义信息的概率不确定性。条件熵 $H(S|X)$ 表示消息可能导致的语义信息的歧义量, $H(X|S)$ 则可用于衡量在传输符号中与语义信息无关的信息冗余量。互信息 $I(X;S) = H(X) - H(X|S)$ 则可用于衡量传输符号中包含的语义信息量。因此, 语义通信理论中的核心问题之一是寻找合适的语义表达以及合适的编码方案, 从而确保用歧义量和冗余量最小的方式传输语义信息。

5.2 语义通信的构成

语义通信一般可通过引入下列组件实现。

5.2.1 语义知识库

语义知识库是语义通信的基石。语义通信的首要前提条件是所有通信参与方(包括发射端和接收端用户)具备相同或者相似背景知识, 其中包括语义实体元素和逻辑及推理关系。具体而言, 语义通信一般需要具备一个或多个公共语义知识库。这些知识库中的实体和逻辑关系可以同时被所有通信参与方接收和理解。除此之外, 不同用户和设备通信中所使用的知识量还和它们的背景、环境、沟通历史等因素紧密相关, 可能存在一定差异。因此, 在语义通信中, 每个用户还可能自行维护和更新一个私有知识库, 用于存储自己或者仅与部分用户共享的私有语义知识和信息。

5.2.2 语义编码器

语义编码器检测并提取源端信号所包含的具体语义内容, 并且压缩和移除与语义无关的信息。编码器还应具备检测源端和终端的知识差异能力, 并可能通过逻辑或其他方法推测出沟通双方共享的知识实体以及关联, 如一个成年人在与小孩子沟通或与成年人沟通时一般会采用不同的词汇库和表达方式, 以确保参与沟通的用户共享相同的知识库。编码器还应当能够处理不同类型的源信号, 如当源信号是一张图片或音频时, 编码器应当首先根据信号源和信号接收端拥有的本地知识识别源图像和音频中的实体, 然后通过一个共同或相似的知识模型识别实体之间可能呈现的关系。实体和关系识别的准确性直接影响信源和信宿之间语义通信性能。换言之, 不同于仅考虑技术问题的传统通信理论。在语义通信中, 通信性能不再主要依赖频谱、能耗等资源数量及接收信号的信噪比等技术指标来衡量, 而需要综合考虑计算量、存

储量和模型精度与标签数据集大小等多种因素对语义的识别、传输和恢复等产生的影响。这些都是仅通过简单延伸传统 Shannon 公式难以表征和分析的。

5.2.3 语义译码器

语义译码器对信源发送的信息进行解释, 并将接收的信号恢复成接收端用户可理解的形式。译码器还需要评估接收端用户的满意程度, 从而判断接收的语义信息是否正确。译码器还可将相关信息(如评估得分)反馈给编码器, 在未来通信中进一步完善实体和关系识别模型。

5.2.4 语义噪声

语义噪声是语义通信过程中引入的可能引起语义信息错误识别和解释的噪声。它可以在编码、传输和译码过程中产生。在编码过程中, 语义噪声可能由于信源信号中实体和它们之间关系被错误地识别导致。信道衰落和噪声也可能引起数据丢失并导致语义失真。在解码过程中, 语义信息接收端对与语义信息的错误解释和用户的误解, 也可能导致语义噪声的产生。

5.3 点对点语义通信的局限

尽管语义通信具有较大的潜力, 但是其发展仍然面临如下挑战。

5.3.1 知识难共享

语义通信的首要前提条件是所有通信参与方(发射端和接收端)能够共享一个或多个普适性的语义知识库, 它一般包含 3 个层次的知识库共享。首先, 发射端和接收端需要拥有一致的知识实体和关系库。其次, 在发射端和接收端拥有的知识库不完全一致的情况下, 发射端和接收端沟通的内容应该在双方共有知识范围内。换言之, 发射端和接收端应当需要具备知识和背景协同以及差异识别能力(如成年人在感知到沟通对象是婴儿或小孩时可自动调整或缩小所使用的知识库)。最后, 当发射端和接收端检测到未知语义知识实体和关系时, 他们之间一般还需要具备知识库协同更新能力。上述 3 个层次知识共享均不可能由单一用户独立更新并维护一个独有的知识库实现, 而需要所有通信参与方共同维护和更新语义知识的相关信息。例如, 即便是对于拥有强大信息处理和记忆能力的人类而言, 知识积累和识别能力也需要花费数十年的时间学习、探索和练习, 而对于通常仅有几年甚至更短使用寿命的电子设备和机器, 每个设备单独维护

和更新知识库不仅需要耗费极其庞大的通信和存储开销，并且知识积累和数据采集也需要耗费大量时间。这些因素共同造成了语义通信难以推广的现状。

5.3.2 语境难感知且难识别

与传统以 0 和 1 为主的数字信号相比，语义信号可能包含的信息和能够表达的内容更加丰富。不同地域和时间用户可能传达与理解的内容也会受各种复杂因素影响，如消息传播者和接收者通信时的性格与情绪、通信环境、与周围用户和环境交互的历史及语义的上下文等。从另一个角度讲，语义通信发生的环境及用户的社交与沟通历史能够有效帮助用户更好地识别语义，减少语义噪声。例如，在教室里老师与学生之间沟通的语义信息有极大可能局限于课堂知识，而在马路或者十字路口无人驾驶车和有人驾车之间沟通的语义信息极大可能聚焦于车辆驾驶行为与交通情况。因此，仅考虑点对点的语义通信的内容而忽视周围用户和通信网络环境影响会极大地限制语义通信中的知识识别和处理效率。

5.3.3 资源需求难满足

语义通信性能与语义编码器和译码器的语义识别和处理准确率密切相关。具体而言，语义通信需要从不同种类的信号源，包括图像、语音、文本中提取用户所需的语义信息，如信号中所包含的物体的某些特征以及不同物体间呈现出的关系。尽管近些年来，由于人工智能技术的快速发展，数据分类、图像与音频信号中的物体识别及逻辑与相关性推理等方面的性能都得到了显著提高，然而大量人工智能算法，尤其是大量基于深度学习的算法，过度依赖高强度算力和高质量人工标记数据集训练出的高性能模型。此外，目前主流的语义通信工作

通常强调可利用先验知识降低通信成本并提高语义传递成功率，却忽略了识别、提取和解释语义信息需要消耗的计算与存储资源开销^[20]。最新发布的报告显示，最先进的人工智能算法耗费的资源在过去的几年中呈持续快速增长趋势，成本每几个月就会翻倍，仅在 2012—2017 年期间，先进的人工智能算法所需资源成本就增加了 300 000 倍^[21]。这些人工智能算法所需要的计算和存储资源量远远超出目前主流终端设备的计算和存储能力，因此进一步制约了点对点语义通信的适用场景。采用一台高性能 GPU 服务器（TITAN X GP102 GPU 和 Intel i9-9900K CPU@3.6GHz）训练 3 种目前主流图像数据集和简单卷积神经网络训练模型精度与耗时如表 1 所示。显然，即便使用高性能 GPU 服务器和标准数据集，训练简单的目标识别模型也需要耗费数分钟，所花费的时间已经远远超出了直接采用目前主流无线通信技术传输高性能图像数据的耗时。

5.3.4 模型协同和隐私安全间的矛盾

语义通信中的发射端和接收端需要持续不断地交换各自获取和感知的知识信息，同时协调和更新各自的知识库，这可能会引起本地知识/隐私数据潜在的泄露风险。此外，持续的信息上传会带来高额的数据传输开销。因此，在语义通信系统中不应采用集中式的框架。然而简单直接采用分布式架构对发射端和接收端之间进行知识协同时，又存在协同通信数据量较大且数据隐私难以保护等问题。

6 语义通信网络组成和范例

6.1 语义通信网络的组成

解决点对点语义通信局限性的有效方法之一是引入具有知识、信息和资源共享能力的语义通信网络。具体而言，本文提出一种基于知识共享和资

表 1 训练 3 种目前主流图像数据集和简单卷积神经网络训练模型精度与耗时

数据集名称	网络模型	本地存储/KB	有标签数据量/个	目标精度	模型训练耗时/s
MNIST	2 层卷积层	46 023	48 000	98.8%	344
	1 层全连接层				
Fashion-MNIST	2 层卷积层	46 092	48 000	87.0%	369
	1 层全连接层				
CIFAR-10	2 层卷积层	121 475	40 000	62.0%	151
	2 层全连接层				
	1 层线性变换层				

源融合的语义通信网络架构，面向万物智联的语义通信网络架构示意图如图 1 所示，该架构有望解决点对点语义通信中的局限性问题。所提架构包含如下的基本单元。

6.1.1 面向网络的语义表征

语义通信网络需要解决的首要问题是知识和语义表征问题。语义通信网络的知识表征应当具备 3 个基本要求：1) 易于搜索，易于实现计算操作；2) 易于添加、删除、更新知识实体和实体间的各种关系；3) 节省存储空间。知识图谱 (KG, knowledge graph) 是一种较易理解和操作的语义形式，可通过图的方式表达接近人类认知世界的各种复杂实体与关系形式，具备良好的组织、管理和理解互联网海量信息的能力。知识图谱是一种由节点 (知识实体, Entity) 与边 (关系, Relation) 组成的多重图结构，一般采用三元组 (主语, 谓语, 宾语) 或 (h, r, t) 表示，如爱因斯坦获得了诺贝尔奖可以采用 $(h=$ “爱因斯坦”， $r=$ “获得”， $t=$ “诺贝尔奖”) 表示^[2]。知识图谱不仅具有易操作和节省空间特点，还可以使用包括图卷积在内的大量人工智能领域的最新工作对不同知识图结构处理、压缩和优化。此外，目前已经有谷歌、WordNet 和 DBpedia 等大量知识图谱可供使用。

6.1.2 边缘计算和存储服务单元

由于识别和处理语义信息所耗费的计算和存储资源远超出单个智能终端所具备的能力，因此，语义通信网络应当充分利用外部的计算和存储资源，并在海量用户之间实现多种资源的融合与共享。在大量用户周围部署边缘计算服务器和存储设备卸载用户终端的语义识别和处理任务是一种有效解决方法。具体而言，用户终端可以实时地将语义编码和解码任务卸载到与其最邻近的边缘服务器上。边缘服务器在收到用户请求后快速执行计算、存储以及 AI 模型训练等任务并将结果及时反馈用户。同时，两个或多个边缘服务器在训练共享知识和语义识别及编解码模型时也可以相互协作，如采用联邦学习和分布式计算等方法，从而进一步减少每个边缘服务器的存储和计算负载，降低单个服务器或用户对标记训练数据集的要求，并提高资源利用率。

6.1.3 全局知识与模型库

全局知识与模型库包含普适性的知识实体 (如常识性的单词和事实) 与不同实体之间的关系。全局知识主要由云计算中心或者分布式边缘计算服务器维护，并可供不同的用户查询和调用。全局语义模型则包括从各种常用的源信号形式 (如文本、

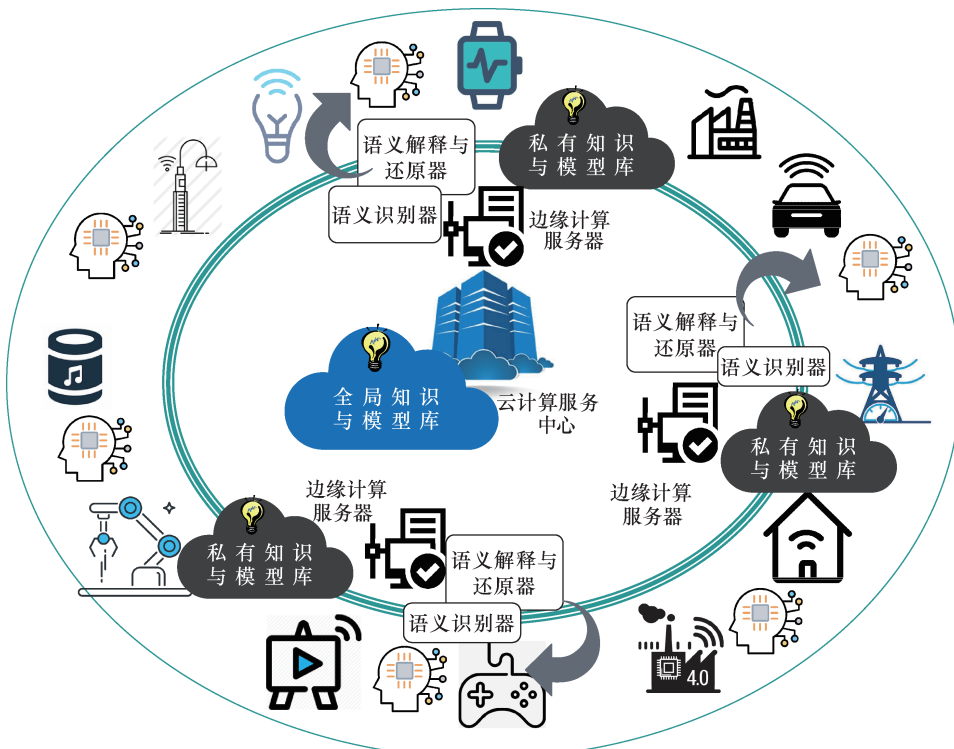


图 1 面向万物智联的语义通信网络架构示意图

图像、语音等)中识别和提取语义信息的主要模型和方法。可通过包括众包、多模态智群感知等在内的多种方式将不同用户在不同背景、语言和场景下积累的普适性知识和语义信息传输给云计算中心或者边缘计算服务器,并将这些知识和信息融入语义识别和认知模型,从而保证知识和语义突破时间和空间的局限,扩大语义通信适用场景。

6.1.4 私有知识与模型库

不同用户可能拥有私有信息,由于语义信息的敏感性,私有信息一般不应上传给其他用户。用户可将私有知识和模型存储在指定的边缘计算服务器,也可以自行设计并训练可识别私有语义信息的模型,并将模型或者参数分享给边缘计算服务器或者其他用户。

6.1.5 用户终端

用户终端可以是低成本的物联网或可穿戴设备,也可以是具备一定计算和存储资源的智慧终端(如手机或智慧屏幕)。如果是后者,用户端可以存储部分简单且经过预训练的语义识别模型,在此情况下,用户仅在所识别出的知识实体精度较低或者无法识别等情况下将语义信息处理任务卸载到边缘计算服务器和云计算中心。

6.1.6 语义识别器

语义识别器包含已训练好的语义知识识别模型(如图像、文本和音频信号中的知识实体识别模型等)。识别器可以集成在边缘计算服务器和存储服务单元内,也可以由第三方物体分类与识别服务供应商(如目前新兴的人工智能服务和模型供应商)提供。

6.1.7 语义解释与还原器

语义解释与还原器主要包含能够还原语义信息并按照接收或发射端用户期望的形式呈现的方法和模型,如可以通过接收端接收的语义信息还原文本、图像、音频甚至全息立体投影呈现给接收端用户。语义解释与还原器还可以根据用户需求和偏好调整和更改语义接收和呈现的形式。

6.2 基于联邦边缘智能的语义通信组网范例

本节提出一种基于联邦边缘智能的语义通信网络架构^[22-24],该架构有望解决前文提到的语义通信局限性问题,联邦学习是一种新兴的分布式机器学习方法。与传统数据集中式的机器学习方法不同,联邦学习可保证多个数据拥有方之间在不泄露本地数据及相关信息的前提下,仅通过交换不泄露

隐私的中间结果实现合作训练和建模。本文提出的语义通信网络架构主要包括如下元素:海量语义通信参与者(用户)、边缘计算服务器和全局知识与语义识别模型协同器。边缘服务器是部署在用户附近的高性能计算服务器,可以执行计算、存储以及AI模型训练等任务。为了降低单个用户部署计算和存储资源的成本,在该结构中,每个用户都可将语义编码和译码等需要耗费大量资源的任务卸载到与其最邻近的边缘服务器。两个或多个边缘服务器可以采用联邦学习方式训练语义和知识识别的机器学习模型,包括全局或局部的语义识别模型。本文采用知识图谱作为语义的表征方法。与Shannon信息论主要采用0和1的离散数字信号刻画信息不同,对于知识图谱来说,用户表达的信息或事件可能是一个存在于知识图谱的子图。因此,每一个可能出现的知识图谱的子图都可以看作一个随机变量,所有可能的子图集合可以看作符号的集合。每个边缘计算服务器在收到用户请求时,首先搜索存储在本地已训练好的知识实体和关系识别模型。当用户上传数据无法被所有本地模型准确地识别时,则开始训练新的语义识别模型。为进一步提高语义知识模型搜索速度,每个用户都应在执行编解码过程之前感知周围环境、通信时间和特定场景以及与通信含义相关的其他信息,这些信息将随源信号一起上传到边缘计算服务器用于分析语义通信的具体场景,从而减少语义编码器和译码器的模型搜索空间。为了协调多个边缘服务器之间的知识建模并保护本地语义数据避免泄露,采用基于联邦学习的分布式模型协同框架,其中每个边缘服务器不需要将其本地数据公布给其他服务器,而只能使用模型训练的中间参数,如使用梯度或者AI模型的参数进行协同,这样能够在保证数据隐私的同时保证模型训练质量,同时利于进一步减少单个边缘服务器的存储和计算负载^[24]。

为了验证所提模型在降低资源损耗和提高通信传输效率等方面的性能,给出如下的仿真结果。首先,评估当允许多个边缘计算服务器采用联邦学习方式协同训练共享的知识(数字、服饰和物体等)识别模型时,所提架构在计算耗时、单个边缘计算服务器所需的存储量和标签数据集的减少量等方面的性能指标。具体地,采用联邦智能边缘训练与表1相同的实体识别模型时单个边缘计算服务器所需的存储量和计算耗时如表2所示。可以观察到,

表 2 采用联邦智能边缘训练与表 1 相同的实体识别模型时单个边缘计算服务器所需的存储量和计算耗时

数据集	目标精度	网络模型	协同边缘服务器数/个	本地存储/KB	所需标签数据量/个	模型训练耗时/s
CIFAR-10	62.0%	2 层卷积层	1	121 475	40 000	151
		2 层全连接层	4	30 551	10 000	62
		1 层线性变换层	8	15 397	5 000	44
MNIST	98.8%	2 层卷积层	1	46 023	48 000	344
		1 层全连接层	4	11 570	12 000	126
			8	5 828	6 000	80
FashionMNIST	87.0%	2 层卷积层	1	46 092	48 000	369
		1 层全连接层	4	11 688	12 000	133
			8	5 887	6 000	121

采用多个边缘计算服务器能够有效地降低模型训练所需时间和每个边缘计算服务器所需存储的带标签数据量。

如前所述，采用知识图谱表征语义知识库的另一个好处是可以利用知识实体之间相关性压缩传输的语义信息。本文采用 3 个主要的知识图谱数据集，包括学术论文引文数据集（Cora）、美国亚马逊网站客户购物数据集（Amazon Co-Purchase）和微软学术图谱中计算机科学领域论文作者数据集（Coauthor CS），模拟在语义通信中能否利用知识图谱相关性压缩传输的数据量，采用半监督图卷积神经网络进行知识图谱压缩时的压缩率与语义信息恢复错误率间关系如图 2 所示，采用半监督图卷积神经网络在传输不同数量的标签数据时的语义信息恢复错误率如表 3 所示。假设发射端仅传输一部分知识实体类别信息，接收端在接收到发射端发送的部分知识实体标签数据集后可以采用半监督图卷积神经网络模型^[25]，恢复出全部知识实体类别信息。选取数据集中部分节点的类别、图的完整邻接矩阵和节点的特征向量进行数据传输，其中图的完整邻接矩阵和节点特征向量以稀疏矩阵形式进行存储，所以占用的数据量很小。所有数据集均采用具有两个图卷积层的网络结构处理，每层通过聚合目标节点和邻居节点特征向量得到新的特征向量作为输出，接收端则可利用最终产生的特征向量进行类别预测。使用 Adam 模型进行训练，训练迭代次数为 200 并且在 10 个连续的迭代次数中损失不再下降即停止训练，两层网络之间传输的隐藏单元数为 16。分析在不同数据压缩率时接收端恢复出的知识图谱的恢复错误概率。可以观察到，由于不同数据集中知识实体间相关性不同，所以在不同的压

缩率下错误率也不相同。具体而言，Coauthor CS 在压缩率为 5.46% 时，仍能够维持 10.91% 的低错误率，但是 Amazon Co-Purchase 在压缩率仅为 50.94% 的时候，错误率就已经高达 14.40% 了，而 Cora 在错误率为 10.93% 时压缩率则为 20.46%。

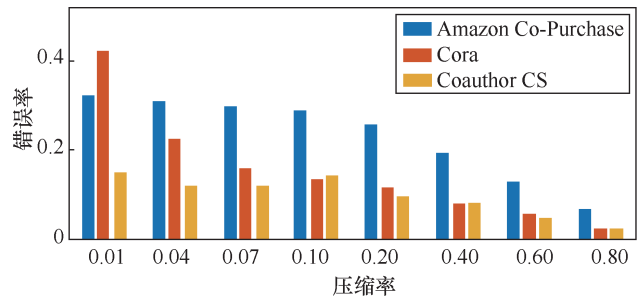


图 2 采用半监督图卷积神经网络进行知识图谱压缩时的压缩率与语义信息恢复错误率间关系

7 语义通信网络的开放问题

7.1 语义多义性解析

语义具有多义性，即相同的句子根据不同场景和内容可能具有不同含义。如英文单词“Catch”在作动词时可以有“理解”的含义，也可能是指“接住”或者“赶上”，在作名词时则可以指“隐患和潜在问题”或者“接住的东西”等。中文字和词的含义则需要根据不同搭配和前后文的具体形式决定，如“包袱”这个词语可以指“布包起来的衣物包裹”，也可以比喻“精神上的负担”。这些不同含义一般可能导致接收端用户在语义理解上出现较大偏差，因此，需要在语义通信编码和译码过程中做出适当解析，以确保接收端能够接收正确的语义。

7.2 多模态协同语义感知与识别

语义可能受不同主客观环境因素影响，如用户

表3 采用半监督图卷积神经网络在传输不同数量的标签数据时的语义信息恢复错误率

数据集	传输标签实体数量							
	节点数量	10	100	190	550	1 150	1 750	2 350
Cora	压缩率	0.52%	3.84%	9.38%	20.46%	42.61%	64.77%	86.93%
	错误率	50.22%	23.15%	15.69%	10.93%	7.31%	4.36%	1.29%
	节点数量	10	100	2 000	5 000	7 000	10 000	13 000
Amazon Co-Purchase	压缩率	0.11%	0.77%	14.58%	36.40%	50.94%	72.76%	94.57%
	错误率	62.88%	31.40%	27.22%	20.34%	14.40%	7.54%	1.97%
	节点数量	10	20	40	1 000	10 000	13 000	17 000
Coauthor CS	压缩率	0.065%	0.12%	0.23%	5.46%	54.56%	70.92%	92.74%
	错误率	75.06%	33.35%	19.95%	10.91%	6.97%	4.37%	0.98%
	节点数量	10	20	40	1 000	10 000	13 000	17 000

情绪、性格与其他通信用户之间社交关系等信息均可能对语义产生影响。单纯依靠单个或者一类传感器(如摄像头)采集的图像信息难以全面识别用户感兴趣的语义信息。因此,需要通过综合分析和融合不同方面、种类和形态的感知数据,如综合分析通信参与用户之间的社交网络信息、视觉图像信息、性格数据等多模态信息数据,提取可能对语义产生影响的多方位因素综合分析,将有望提升语义通信的识别效率和精度。

7.3 环境变化的快速感知、识别和预测

语义可能随环境变化而发生快速变化,如天气变化或者通信参与用户社交状态和注意力变化等均可能导致语义信息发生变化^[26]。部分环境变化具有一定规律性和可预测性,如果能够充分利用这些性质对可能影响语义的环境因素和时变特征进行快速感知、识别和预测,并提前开展语义的预分析与模型的预训练,将能够进一步提高语义识别速度,同时可以在时间和空间维度上实现不同资源的优化适配与合理调度。

7.4 语义知识库的自学习与自更新架构

由于语义知识可能随着人类社会发展而变化,语义知识库需要具有自学习和自更新功能^[1]。具体而言,当新兴语义知识实体(如新名词或者新术语)在某个局部或者平台出现时,语义知识库需要根据该实体使用和引用情况识别和分析其可能具备的含义和使用场景,最后将该实体与其他已有实体间的关系信息添加到知识库。同时,语义知识库还应当能够持续跟踪该实体后续的使用和引用情况并进行实时更新。

7.5 语义通信网络安全性

语义通信网络对安全性提出了全新要求。所有上传到语义知识库的信息应当具有完善的审查机

制,防止语义知识库被恶意用户篡改。除此之外,每个用户的私有语义信息和全局或局部共享的语义信息之间应当具有安全可靠的存储和调用机制,以防止用户隐私信息泄露。相信随着语义通信网络发展,会有更多新兴的安全问题暴露出来,提前预测、建模和分析未来可能出现的新兴语义通信网络安全隐患对语义通信技术的发展和普及具有至关重要的意义。

8 结束语

本文从面向未来万物智联网络出发,首先,介绍语义通信与万物智联网络的关系,通过分析通信的内涵,提出语义通信将有望成为未来通信技术发展的新范式这一技术发展趋势。随后,介绍了语义通信的基本模型和组成。通过分析点对点语义通信的局限性,提出以网络知识共享和资源融合为基础的语义通信网络将更适合成为万物智联网络的基础范式。通过分析语义通信网络的基本组成,提出了一种基于联邦边缘智能的语义通信网络架构范例。仿真结果显示语义通信网络可以在保证数据安全的同时大幅度降低语义通信资源需求,并提高通信效率。最后,介绍了语义通信网络未来发展的开放性问题的。

参考文献:

- [1] XIAO Y, SHI G M, LI Y Y, et al. Toward self-learning edge intelligence in 6G[J]. IEEE Communications Magazine, 2020, 58(12): 34-40.
- [2] ITU-T. Architectural framework for machine learning in future net-works including IMT-2020: Recommendation ITU-T Y.3172[S]. 2019.
- [3] 石光明,李莹玉,谢雪梅.语义通讯:智能时代的产物[J].模式识别与人工智能,2018,31(1):91-99.
SHI G M, LI Y Y, XIE X M. Semantic communications: outcome of the intelligence era[J]. Pattern Recognition and Artificial Intelligence,

- 2018, 31(1): 91-99.
- [4] XIAO Y, HIRZALLAH M, KRUNZ M. Distributed resource allocation for network slicing over licensed and unlicensed bands[J]. *IEEE Journal on Selected Areas in Communications*, 2018, 36(10): 2260-2274.
- [5] XIAO Y, KRUNZ M. Distributed optimization for energy-efficient fog computing in the tactile Internet[J]. *IEEE Journal on Selected Areas in Communications*, 2018, 36(11): 2390-2400.
- [6] HOCKETT C F, SHANNON C L, WEAVER W. The mathematical theory of communication[J]. *Language*, 1953, 29(1): 69.
- [7] COVER T M, THOMAS J A. Elements of information theory[M]. New York, USA: John Wiley & Sons, Inc., 1991.
- [8] SHANNON C E. A mathematical theory of communication[J]. *The Bell System Technical Journal*, 1948, 27(3): 379-423.
- [9] CARNAP R, BAR-HILLEL Y. An outline of a theory of semantic information: RLE (Research Laboratory of Electronics) Technical Reports 247[R]. Massachusetts Institute of Technology, Cambridge MA, 1952.
- [10] FLORIDI L. Outline of a theory of strongly semantic information[J]. *Minds and Machines*, 2004, 14(2): 197-221.
- [11] BAO J, BASU P, DEAN M K, et al. Towards a theory of semantic communication[C]//Proceedings of 2011 IEEE Network Science Workshop. Piscataway: IEEE Press, 2011: 110-117.
- [12] 石光明. 语义通信变革现代通信模式[J]. 战略前沿技术, 2020. SHI G M. Semantic communication transforms modern communication patterns[J]. *Strategic Frontier Technology*, 2020.
- [13] SHI G M, XIAO Y, LI Y Y, et al. From semantic communication to semantic-aware networking: model, architecture, and open problems[J]. arXiv: 2012.15405, 2020.
- [14] 钟义信, 张瑞. 信息生态学与语义信息论[J]. 图书情报知识, 2017(6): 4-11. ZHONG Y X, ZHANG R. Information ecology and semantic information theory[J]. *Documentation, Information & Knowledge*, 2017(6): 4-11.
- [15] 钟义信. 信息理论的新阶段: 语义信息论[Z]. 2016. ZHONG Y X. The new stage for information theory: semantic information theory[Z]. 2016.
- [16] 张平. 移动通信堆叠处理模式难以为继 6G 亟需原创理论创新[Z]. 2020. ZHANG P. Challenges in traditional stacking-based processing mode urging original theoretical innovation in 6G[Z]. 2020.
- [17] JUBA B. Universal semantic communication[M]. Berlin: Springer, 2011.
- [18] JUBA B, SUDAN M. Universal semantic communication[C]//Proceedings of the 40th annual ACM symposium on Theory of computing. New York: ACM Press, 2008.
- [19] B JUBA, M SUDAN. Universal semantic communication II: a theory of goal-oriented communication[C]//Electronic Colloquium on Computational Complexity (ECCC) TR08-095. [S.l.: s.n.], 2008.
- [20] GÜLER B, YENER A, SWAMI A. The semantic communication game[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2018, 4(4): 787-802.
- [21] SCHWARTZ R, DODGE J, SMITH N A, et al. Green AI[J]. *Communications of the ACM*, 2020, 63(12): 54-63.
- [22] XIAO Y, LI Y, SHI G, et al. Optimizing resource-efficiency for federated edge intelligence in IoT networks[C]//Proceedings of International Conference on Wireless Communications and Signal Processing. [S.l.: s.n.], 2020: 86-92.
- [23] ZHOU Z, CHEN X, LI E, et al. Edge intelligence: paving the last Mile of artificial intelligence with edge computing[J]. *Proceedings of the IEEE*, 2019, 107(8): 1738-1762.
- [24] MCMAHAN B, MOORE E, RAMAGE D, et al. Communica-

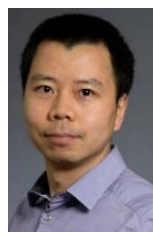
tion-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. PMLR, 2017: 1273-1282.

- [25] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[EB]. 2016
- [26] XIAO Y, KRUNZ M. Dynamic network slicing for scalable fog computing systems with energy harvesting[J]. *IEEE Journal on Selected Areas in Communications*, 2018, 36(12): 2640-2654.

[作者简介]



石光明 (1965-)，男，博士，西安电子科技大学教授、博士生导师，人工智能学院领军教授，教育部长江学者特聘教授，享受国务院政府特殊津贴专家，IEEE/IET Fellow，中国电子学会会士，人工智能学会高级会员，2017 年以第一完成人荣获国家自然科学奖二等奖 1 项，主要研究方向为语义通信、类脑感知、压缩感知、计算成像、机器学习、脑电信息处理与脑启发智能技术等。



肖泳 (1980-)，男，博士，华中科技大学教授，IMT-2030 (6G) 推进组网络智能方向副组长，5G 联创行业应用开发实验室副主任，IEEE 高级会员，中国通信学会高级会员，IEEE *Transactions on Mobile Computing* 副编辑，主要研究方向为网络人工智能、边缘计算、通信网络博弈理论等。



李莹玉 (1991-)，女，博士，华中科技大学电子信息与通信学院在站博士后，主要研究方向为物联网、网络大数据分析、分布式优化理论等。



高大化 (1979-)，男，博士，西安电子科技大学人工智能学院教授、博士生导师，陕西省“青年科技新星”，主持国家重点研发计划、国家自然科学基金等项目 10 余项，发表智能计算成像相关论文 30 余篇，授权国际、国内专利 10 余项，主要研究方向为智能计算成像、智能信息处理等。



谢雪梅 (1967-)，女，博士，西安电子科技大学人工智能学院教授、博士生导师，IEEE 高级会员，入选教育部“新世纪优秀人才支持计划”，获得第十八届中国科协年会全国科技工作者创新创业大赛陕西赛区银奖，在国际期刊和会议上发表了 100 多篇学术论文，主要研究方向为人类动作识别、目标检测、场景理解、视频分析、深度学习和特征表示等。