# 面向6G边缘智能的语义数据源寻

Kaibin Huang、Qiao Lan、Zhiyan Liu 和 Lin Yang

arXiv:2301.00403v1 [cs.IT] 1 Jan 2023

摘要—**作为6G网络的一项新功能，边缘智能是指在网络边缘无处不在地部署机器学习和人工智能（AI）算法，以支持从感知到自动驾驶等众多新兴应用。为支持包括感知、边缘学习和边缘推理在内的相关应用场景，均需要通过空中传输高维数据或人工智能模型。为克服这一瓶颈，我们提出了一种新型的语义数据采集（SEMDAS）框架，用于定位语义匹配的数据源，从而高效实现边缘智能操作。该综合框架包括全新的架构、协议、语义匹配技术，以及面向任务的无线技术设计原则。作为SEMDAS的关键组成部分，我们讨论了一组基于机器学习的语义匹配技术，旨在适配不同的边缘智能应用场景。此外，针对面向任务的无线技术设计，我们讨论了SEMDAS系统中的不同权衡，提出了语义与信道联合匹配这一全新概念，并指出了若干研究机会。SEMDAS框架不仅克服了上述通信瓶颈，还解决了其他网络问题，包括远距离传输、稀疏连接、高速移动性、链路中断和安全问题。此外，我们使用真实数据集开展实验并展示了实验结果，以证明SEMDAS带来的性能增益。**

## I. 引言

预计第六代移动网络（6G）将原生支持人工智能（AI），其特点是在网络边缘普遍部署机器学习和人工智能算法 [1]。另一方面，数据已取代燃料，成为世界上最宝贵的资源 [2]。移动数据正以指数级增长的速度生成，预计将增长三倍，到 2028 [3]年达到每月324艾字节。6G边缘智能将提供一个持续提炼人工智能的平台，以支持从感知、自动驾驶到工业自动化等众多万物互联（IoE）应用。与“云端人工智能”相比，边缘智能具有高效处理移动数据、保护用户隐私、减少网络流量以及提供超低时延接入等优势 [4]。实施边缘智能面临的主要挑战在于，对高维数据或人工智能模型参数进行无线传输会造成通信瓶颈。为克服这一瓶颈，本文提出一种名为语义数据源寻（SEMDAS）的新型框架，用于定位语义匹配数据源，从而高效支持边缘智能操作。

### *A. 5G*连接中心网络

5G 的关键创新在于提供一种基础设施，以支持异构类型的服务和应用。具体而言，为满足速率、可靠性和时延方面的不同要求，定义了三种连接类型：超可靠低时延通信、海量机器类通信和增强型移动宽带。5G 通信技术的设计很大程度上采用了源自香农信息论的速率中心方法。在这种方法中，数据基本上被视为一串比特，需要可靠、快速地从源传输到目的地。正如沃伦·韦弗在其 1953年 的论文中所指出的 [5], 通信应超越单纯解决这一技术问题，转而处理语义问题，即准确传达消息含义，以及消息在目的地被有效用于执行特定任务。有人会认为，许多移动应用确实具备语义感知能力，例如推荐和广告。然而，这些应用只是网络的附加组件，仅与其应用层交互。换言之，它们与部署无线技术的无线接入层相互独立。传统的计算与通信分离方法被认为在端到端（E2E）性能方面并非最优。因此，现有的语义无关技术缺乏应对呈指数增长的移动数据和边缘设备数量、实现快于人类反应速度（即 0.1 毫秒）时延所需的最高效率。因此，6G 研究人员倡导采用新的计算与通信融合方法，在设计下一代无线技术时紧密结合通信、计算、感知和控制 [1]。

### *B. 6G*语义通信

作为一种 6G 范式，语义通信 （SemCom）是指计算通信融合设计，其优化目标是语义准确性这一端到端性能指标，或任务有效性的重叠度指标。与传统的不透明数据传输相比，语义感知可以通过避免传输缺乏相关性的信息来减少通信开销，并提高无线资源利用效率（例如，根据数据内容在资源分配时对数据包进行优先级排序）。SemCom 系统中的两项基本操作是语义编码（在保留消息语义或任务效用的同时对消息进行压缩）和信道编码（用于在存在信道失真的情况下确保可靠性）。在一种称为联合信源信道编码的常见方法中，这两项操作由独立的神经网络模型实现，并针对边缘推理或感知等端到端任务进行联合训练 [6]。为了在分层网络架构上实现，SemCom 涉及对顶层应用层和底层无线接入层的深度耦合

K. Huang、Q. Lan 和 Z. Liu 任职于香港大学电机电子工程系，中国香港特别行政区。L. Yang 任职于中国华为诺亚方舟实验室。联系人：K. Huang（huangkb@eee.hku.hk）。

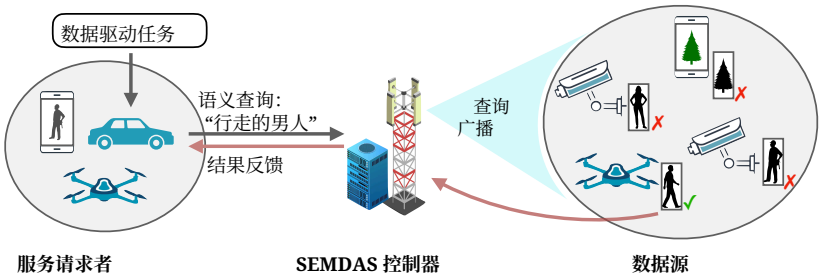


图1。SEMDAS网络架构。

[7],[8]。为说明其优势，考虑传输如下消息：“爱因斯坦的儿子 **** ****** ********，曾是**********大学********* **********专业的教授。”其中错误的字母用“*”标记。该消息在物理层可能会因不可靠而被丢弃，但在应用层借助知识图谱可以将其可靠地恢复为：“爱因斯坦的儿子汉斯·阿尔伯特·爱因斯坦，曾是加州大学伯克利分校水利工程专业的教授。”边缘智能的兴起为 SemCom 提供了一个平台，其中包括人工智能算法、知识图谱和数据分析在内的强大工具能够实现系统运行的自动化和简化。此外，鉴于具备抗扰动能力的人工智能赋能接收机，即使传输的源数据样本特征数量减少，推理性能仍可平稳下降，这也使接收到的消息即使在丢包情况下仍然具有实用价值。

### *C.* 面向*6G*边缘智能的语义数据采集方案

为了在语义通信平台上高效实现边缘智能，我们提出了一种名为语义数据采集（SEMDAS）的新型框架。SEMDAS旨在从众多数据源中定位能够提供语义相关数据的子集，以支持通信高效的边缘学习或推理。为此，SEMDAS协议本质上要求服务请求方向控制器传输一个查询，该查询描述特定的边缘智能任务；控制器则搜索匹配的数据源。随后，匹配的数据源与请求方建立连接，以提供计算服务或其数据。作为SEMDAS的关键组成部分，我们针对边缘智能的三种代表性应用场景——物联网感知、边缘学习和推理——提出了一组语义匹配技术。这些技术建立在经过适当选择的现有学习算法之上，包括语义嵌入、不确定性评估、分布匹配和自编码器。

所提出的 SEMDAS框架具有三大优势。首先，SEMDAS 通过避免传输不必要的语义无关数据，克服了边缘智能的通信瓶颈。其次，无需端到端连接，SEMDAS 利用多个语义相似源的存在，解决包括远距离传输、稀疏连接、高速移动性和链路中断在内的网络问题。最后，数据源验证——SEMDAS 控制器通过语义检查进行数据源验证，提供了一种确保安全和数据完整性的机制。

我们进一步提出了设计面向任务的 SEMDAS 无线技术的新原则。其中一项原则是基于查询大小、数据开销、隐私和语义匹配准确率之间的权衡进行系统优化。另一项原则是联合进行语义-信道匹配。我们讨论了这些新的权衡如何促成面向任务的无线技术设计，包括多址接入、空中计算、无线电资源管理和波束成形。

## II. 语义数据采集概述

尽管 SEMDAS 适用于更广泛的应用场景，我们在此聚焦于边缘智能领域。下文各小节将分别介绍 SEMDAS 的网络架构、协议和优势。

### *A. SEMDAS* 网络架构

如图 1 所示，该架构由以下关键组件构成。

- **服务请求方：**该节点（设备或服务器）执行一项数据驱动任务。为此，它会向 SEMDAS 控制器发送一个包含描述符的查询，以请求有助于有效执行任务的语义匹配数据。
- **SEMDAS 控制器：**该节点协调 SEMDAS 流程，作为数据源与请求方之间的接口，实现安全功能并管理移动性。
- **语义数据源：**数据源可以是边缘设备或服务器，用于存储和提供感知数据/用户数据、多媒体文件、文档或人工智能模型。语义数据是根据其语义（即内容和效用）分类的数据。单个数据样本中嵌入着不同类型的语义信息，例如同一图像中的人脸和行为、建筑物、天气、上下文和位置。对于特定任务，只有选定的信息有用（例如，监控所需的人脸）。语义数据集由其语义标识，而不受物理位置、生成机制和通信方式的影响。两个数据集之间的语义相似度可根据给定的描述符来衡量，例如“笑脸”“德国牧羊犬”和“手势”。描述符也可以是多媒体对象的形式，

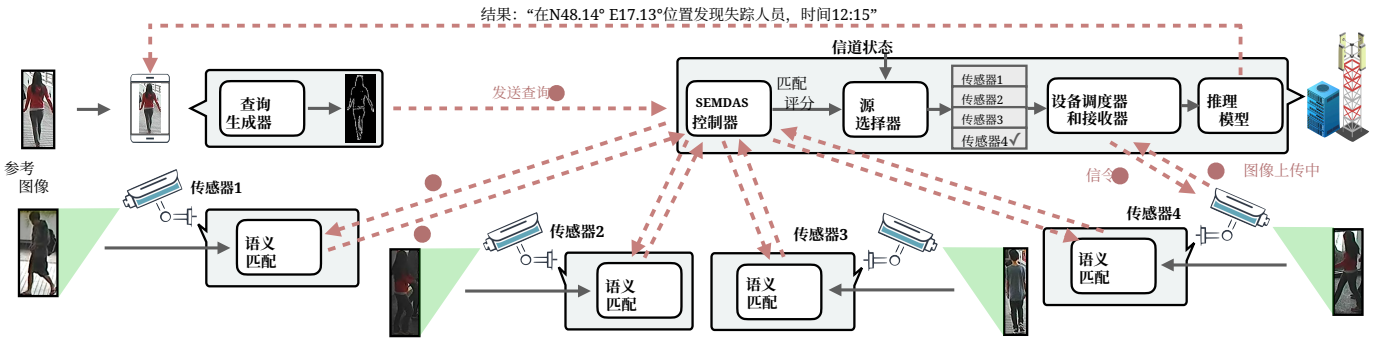


图2. 用于人工智能赋能的物联网感知的 SEMDAS。

例如图像、视频片段和语音信号（如第五节中的失踪人员照片）。从任务有效性的角度来看，任何两个语义相似的数据集在效用上都是相同的。存储语义相似数据的节点都可以向同一个请求方提供数据。

- **语义匹配：**SEMDAS可视为一种面向任务、基于语义的边缘智能“搜索引擎”。人工智能算法用于实时匹配数据与查询之间的语义相似度。具体算法将在下一节讨论。

### *B. SEMDAS*协议

典型的SEMDAS协议包括以下步骤。

- **服务请求：**服务请求方向SEMDAS控制器发送查询（例如文本或多媒体对象）。查询通常是一个低维特征向量，用于描述所需的数据。请求方可以是用户设备（如物联网感知和边缘推理场景中的设备），也可以是边缘学习场景中的服务器。
- **语义源选择：**控制器向网络中的所有数据源广播查询。每个源都会为其数据生成一个键，并使用合适的语义匹配技术（见第III节）将其与查询进行比较。该操作会生成一个匹配分数，并将其反馈给控制器。

- **数据上传：**利用匹配评分，服务器会选择匹配度最高的数据源并上传其数据。数据随后转发至请求节点进行处理。或者，也可以由控制器或服务器执行计算，再将结果下载给用户。

### *C. SEMDAS* 的优势

所提出的 SEMDAS 框架具有以下主要优势。首先，SEMDAS 为边缘智能的数据采集提供了一种可扩展的解决方案。它通过评估数据源之间的语义相似度，避免不必要的传输，从而克服点对点数据传输的瓶颈；同时利用多个语义相似数据源的存在，应对长距离传输、链路中断和站点无法访问等通信问题。其次，通过检查其语义内容，SEMDAS 控制器可验证数据源，从而确保安全性和数据完整性。第三，SEMDAS 有助于移动性管理。与 5G连接中心网络不同，SEMDAS 方法无需端到端连接。移动用户重复发送相同的查询，不同的语义相似数据源均可为其提供服务。无需维持与先前数据源的连接。因此，SEMDAS 有助于应对稀疏连接、高速移动性和链路中断等不利网络状况。

## III. 面向边缘智能的语义匹配技术

### *A.* 人工智能赋能的物联网感知中的语义匹配

物联网感知是 6G 的一项新功能，它利用设备端传感器之间的跨网络协作，构建大规模传感器网络，用于监控、定位、跟踪和事件检测。将多模态感知数据输入边缘人工智能模型，可使传感器具备物体/事件识别和人类行为检测能力。然而，传输高维感知数据会给网络带来沉重负担。基于语义匹配进行传感器选择，可以缓解由此造成的网络拥塞。相关应用和技术如下所述。

考虑两种具有代表性的应用场景。第一种，众包感知是指由一组参与用户所拥有的传感器共同执行感知任务。如图 2 所示，在第 V 节的实验中，我们考虑了一项寻找失踪人员的具体任务。第二种应用场景是网络化感知。网络功能涉及多个设备（例如车辆和机器人）协同完成跟踪、定位和目标识别等感知任务，并由边缘服务器进行协调。网络化感知这一应用场景还可进一步分为两种子场景——对等辅助感知和多视角感知。前者通过使用协助设备收集的匹配数据，解决设备上的传感器退化问题。例如，摄像头故障的自动驾驶汽车可以依靠附近车辆观测道路及周围环境 [9]。另一方面，多视角感知利用服务器汇总不同视角下多个摄像头传感器的二维观测

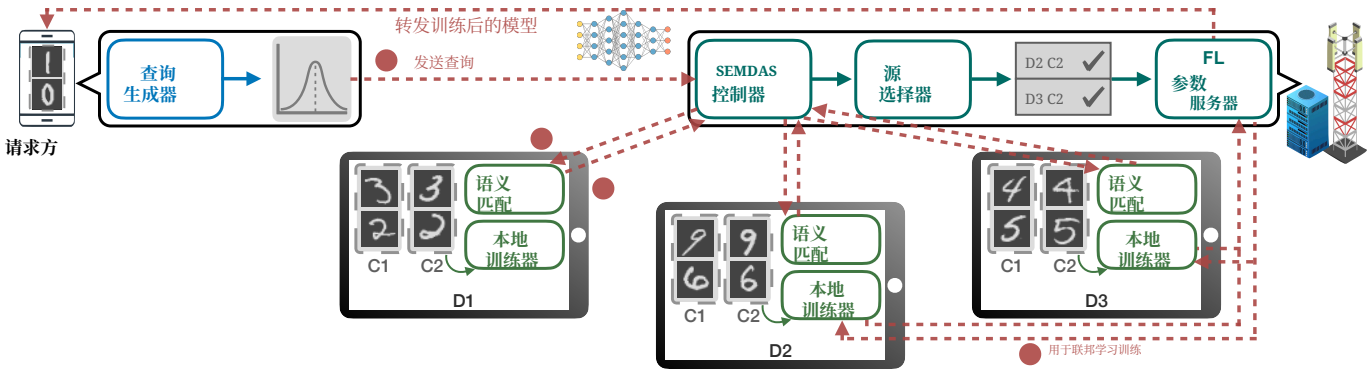


图 3. 用于联邦边缘学习的 SEMDAS。

，以提高感知精度（例如目标识别），或重建三维物体 [10]。例如，在对等辅助感知场景中，查询可以包含由用户的故障传感器拍摄的退化视图图像；在多视角感知场景中，则可以根据野猪图像生成查询，以定位市中心的野生入侵者。

我们提出了一种高效的语义匹配技术，旨在语义空间中找到与查询接近的观测。语义空间是一种特征空间，其中含义相似的样本（例如短语或图像）会聚集在一起，而含义不同的样本则彼此分离。实际上，语义空间是通过投影神经网络模型创建的，该模型根据相关感知任务进行训练，并用于将查询和键投影到空间中 [9]。随后，利用通用注意力机制计算查询-键对的匹配分数；该机制会计算它们在所学习的语义空间中的投影的点积或余弦相似度。例如，Google RankBrain 系统为每天数十亿人使用的 Google 搜索引擎提供支持，它在语义空间中匹配查询和网页，而非进行简单的关键词搜索。该系统推断查询背后的语义意图，并识别涵盖该意图的网页；这可以理解为将两者投影到语义空间中，再依据相似度测量进行筛选。

上述语义匹配技术要实现最佳性能，需要对两个语义编码器进行端到端训练。一个是查询生成器，用于将请求投影为低维查询；另一个是键生成器，用于计算承载感知观测中语义信息的键。

*B.* 边缘学习的语义匹配

边缘学习是指一种边缘智能应用场景：利用分布在无线网络中的移动数据训练人工智能模型。它被视为一项关键的 6G 业务，可提炼出赋能广泛应用所需的智能。主要有两种方法。第一种是集中式边缘学习 (CEEL)，将设备中的数据（或其特征）直接上传至服务器进行模型训练。另一种方法是联邦边缘学习(FEEL)，适用于需要保留数据所有权的情况 [4]。为此 FEEL 使用所谓的参数服务器，将全局模型下载到设备上，利用本地数据集进行更新，然后上传并聚合本地模型，以更新全局模型。该过程采用随机梯度下降这一经典迭代算法，重复执行直至模型收敛。两种方法都面临通信瓶颈，因为它们需要从可能数量众多的设备上传高维数据或模型参数/梯度。SEMDAS 方法可以缓解这一瓶颈（见图 3）。

不同设备上的数据表现出高度异质性。对所有数据不加区分地训练模型可能有害，因为纳入域外数据会损害模型。因此，SEMDAS 的目的在于寻找与学习任务领域相匹配的数据。例如，要学习法英机器翻译，其他语言的数据集就无关紧要。再如，要训练一个能够识别某位著名作家手写文本的模型，并非所有含义相同的手写文本都有用，只有与该作家风格相匹配的文本才有用。因此，在边缘学习的上下文中，语义匹配指的是训练数据与学习任务之间的领域匹配。

下文将讨论两种情况下的领域匹配技术：数据集由标签或描述符表征。具有数据描述的情况相对简单。控制器接收数据源发布的数据描述，以及数据请求发送方发送的查询中的任务描述，即可执行数据-任务匹配。随后，控制器将代表性数据和任务描述符投影到语义空间中，以评估二者的相似性，并据此评估特定数据集与任务之间的匹配分数；所采用的技术与第 III-A 节中的技术相同。对于没有数据集描述的情况，领域匹配更具挑战性。近期，研究人员观察到，同一标签的样本簇中嵌入的不同隐式语义特征，可能导致该簇呈现嵌套子簇分布 [11]。例如，在 MNIST 数据集中，同一标签的手写数字会形成多个子簇，每个子簇对应一种书写风格 [11]。因此，领域匹配可以转化为数据分布与任务之间的匹配。为此，

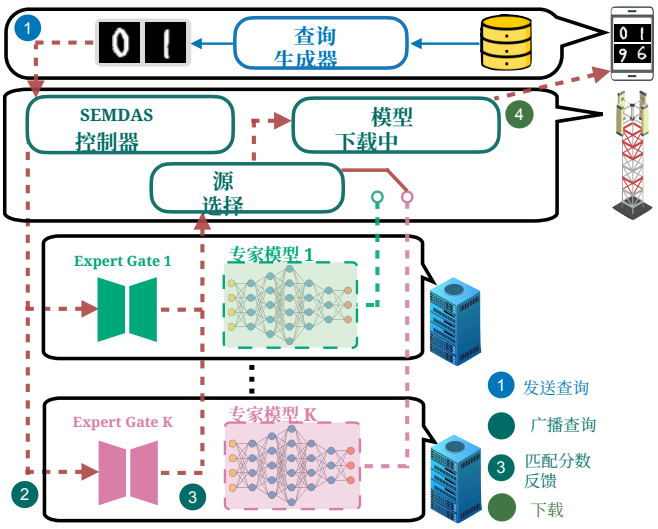


图 4. 用于模型下载的边缘推理 SEMDAS。

查询中会包含目标领域数据的统计信息，或此类数据的代表性小批量。随后，通过评估或近似特定数据集与查询所指定目标分布之间的 *Kullback-Leibler*距离，生成该数据集的匹配分数。在 CEEL 中，控制器会协调选定的设备将其数据上传至请求服务器；在联邦边缘学习中，则协调这些设备参与其中。

### *C.* 边缘推理的语义匹配

边缘推理是6G边缘智能的一项基本操作，着重于向边缘设备高效提供人工智能推理能力。根据是否将计算卸载到边缘服务器，边缘推理可通过两种方式实现，即设备端推理和拆分推理。前者是指按需将人工智能模型下载到设备上进行本地推理，如图4所示。这适用于中小型模型大小，能够提供更快的响应速度和更好的隐私保护。对于后者，计算负载由用户和服务器分担。设备子模型从数据样本中提取中间特征，避免共享原始数据，从而保护数据所有权。随后，服务器接收上传的特征，并将其输入服务器子模型进行预测。通过将密集计算卸载到服务器，拆分推理使资源受限设备能够使用大型深度神经网络模型，而这些模型无法在设备端部署。网络中存储的模型数量实际上近乎无限。这些模型可从许多维度加以区分，例如大小、上下文、任务、性能和拓扑结构 [12]。例如，目标识别模型可以采用卷积拓扑结构或递归拓扑结构，并且可以针对城市、农村或室内环境进行训练。在边缘推理的语境下，数据源指存储特定可共享模型（即其参数和拓扑信息）的节点（设备或服务器）。因此，SEMDAS问题就是寻找模型符合请求方在推理准确率、复杂度和存储方面要求的数据源。

下面讨论边缘推理中的语义匹配技术。首先，对于模型描述可用的情况，所采用的技术与感知和边缘学习中的技术类似，匹配过程会利用语义空间。接下来，我们关注模型描述不可用或不足以进行语义匹配的情况。为此提出的技术要求请求设备生成一个查询，其中包含本地数据（例如传感器观测）的测试样本。然后，SEMDAS 控制器使用该查询，通过以下两种技术之一执行语义匹配。第一种技术是专家网关。假设控制器可以访问一个热门人工智能模型库，这些模型称为专家模型。除了推理模型本身外，每个专家还配备一对轻量级自编码器，包括编码器和解码器。控制器将查询提供给每个专家进行评估。作为响应，专家使用其自编码器从测试样本中提取特征，并生成重构误差 [13]。控制器选择误差最小的专家（即与用户任务最匹配的专家），并下载其模型。第二种技术称为服务器轮询，如果专家模型中没有可匹配的模型，SEMDAS 控制器就会向网络中的所有可用服务器广播用户的查询。每台服务器都使用自己的模型对查询中的测试样本执行推理，并评估生成的置信度分数，以此检验模型的匹配程度。总体评分最高的模型被视为与该查询在语义上匹配。随后，在拆分推理的情况下，将托管该模型的服务器与请求用户关联，以提供推理服务。在模型下载的情况下，则会检索模型并转发给用户。

## IV. SEMDAS 的任务导向通信

SEMDAS 的任务导向通信技术采用端到端性能指标进行设计，例如物联网感知准确率、联邦边缘学习的收敛速度，以及边缘推理的预测误差。本节将通过提出新的设计原则并探讨若干研究机会，介绍这一新领域。

### *A.* 权衡与最优设计

SEMDAS 的无线通信中存在两种独特的权衡。它们有助于优化边缘智能系统和相关技术。

第一个权衡称为查询-数据权衡，源于两种相反的趋势。一方面，缩短查询向量可减少向所有数据源广播所产生的开销。另一方面，查询所包含的信息减少，可能导致语义匹配的准确度降低，进而需要上传更多数据，从而增加上行链路开销。反向趋势同样成立，因此形成了查询开销与数据开销之间的这种权衡。可以基于这一权衡优化 SEMDAS 的无线系统。例如，可以针对某个端到端性能指标，优化带宽或时间的上行链路与下行链路划分

此外，在查询速率约束下，可以设计查询，以最大化数据源处的整体语义匹配程度；或者在给定匹配程度的情况下，最小化数据上传的通信开销。

第二种权衡称为<样式 id='1'>隐私-准确率权衡</样式>。其原因在于，在查询中提供更多细节（例如失踪人员的特征数量），会泄露更多关于请求方的私人信息（例如该人员的身份）或其任务的信息，但同时也能提高语义匹配准确率，进而提升通信效率，反之亦然。这种权衡具有实际意义，因为在许多实际场景中，查询包含私人信息，而将其通过网络广播可能会损害请求方隐私。考虑这种权衡后，就需要设计隐私保护通信方案。具体而言，在对请求方施加隐私约束的情况下，可将查询生成与自适应传输（例如编码、功率控制和波束成形）相结合，以最大化系统效率。

### *B.* 联合语义与信道匹配（*JSCM*）

采用速率中心方法时，仅依据其信道状态选择数据源，可能导致所获取的数据与计算任务不匹配。然而，选择时若只考虑语义匹配，则可能导致通信链路不可靠。因此，为了支持无线系统中的 SEMDAS 应用，选择数据源时应兼顾这两个方面，这称为 JSCM。重新设计传统速率中心无线技术，使其具备 JSCM 特性，将带来许多研究机会。以下将介绍其中几个代表性方向。

- **多址接入：** 基于 JSCM 的分布式数据源选择，可通过分别对其语义匹配分数和信道增益设置阈值，来选择通过多址接入信道上传的数据源。可联合设计这两个阈值，以实现双重目标：获取足够的相关数据，同时控制接入设备的数量，避免频繁发生数据包碰撞或无线电资源不足。

- 空中计算 （**AirComp**）： AirComp 利用多址接入信道的波形叠加特性，在感知和联邦边缘学习（FEEL）用例中分别实现视图以及本地模型／随机梯度的空中聚合 [14]。抑制 AirComp误差的瓶颈在于条件不利的发射机信道。基于 JSCM 的设备选择可通过兼顾抑制 AirComp误差和获取足够的数据，来优化端到端系统性能。

- 无线资源管理 **(RRM)：** 传统的无线资源管理技术，例如子信道分配、功率控制和调度，都是基于速率中心的指标设计的。为了使其面向任务并适用于SEMDAS，这些技术应重新设计，以便不仅根据设备的信道状态，还根据其语义匹配分数来确定优先级，而不是仅根据其信道状态，还根据其语义匹配分数。

- **随机波束成形：** 现有的随机波束成形方案利用多用户信道分集，通过从众多用户中选择具有波束对齐的信道向量和较大信道增益的用户，来最大化数据速率。此类方案可以加以改进，同时利用多个数据源在信道和语义方面的分集。

## V. 实验

### *A.* 实验设计

我们通过一项利用无线连接的监控摄像头网络寻找失踪人员（例如儿童或老年人）的感知应用来展示 SEMDAS。该底层操作在技术上称为 重新识别（ReID），旨在将不同时间拍摄的摄像头视图与同一人员关联起来，该人员由其家属在查询中发送的参考照片指定。匹配摄像头共享的实时位置信息有助于定位该人员。为实现语义匹配，我们采用在广为人知的 CUHK-03 训练集 [15] 上训练的 LightMBN 模型进行语义特征提取，并使用第 III-A 节中讨论的匹配函数。因此，查询照片和每个摄像头视图均被压缩并转换为 3584×1 的特征向量，以用于语义匹配。在通信方面，向传感器广播的查询在每个维度上采用 32 位分辨率；匹配的传感器上传其视图的原始数据（即照片），并确认“找到人员”。我们模拟了一个由 20 个摄像头组成的网络，并从 CUHK-03 数据集中抽取样本来生成查询和摄像头视图。在每次试验中，20 个摄像头视图中有 4 个包含目标身份，其余视图则对应不同人员。实验总共包含 1229 次随机试验。传感器到 SEMDAS 控制器的信道建模为独立同分布的瑞利衰落信道，并设为正交信道，总上行带宽为 5 MHz。称为丢失率的性能指标定义为 ReID 失败的概率，即所有匹配摄像头视图中均不包含目标人员的概率。丢失率根据每次请求的平均上行通信负载（以 Mbit 计）和平均上行通信时延（以毫秒计）进行评估。在实验中，我们考虑以下 JSCM 和基准方案。

- （提出的）JSCM：根据最大化语义匹配分数与通信速率加权和的准则，选出指定数量的匹配传感器，其中二者的权重经数值优化后分别为 1 和 0.09。

- 最佳语义选择 （BSS），依据语义匹配分数最大化准则进行选择。
- 最佳信道选择 （BCS），依据通信速率最大化准则进行选择。

- 上传传感器的随机选择（RS）。

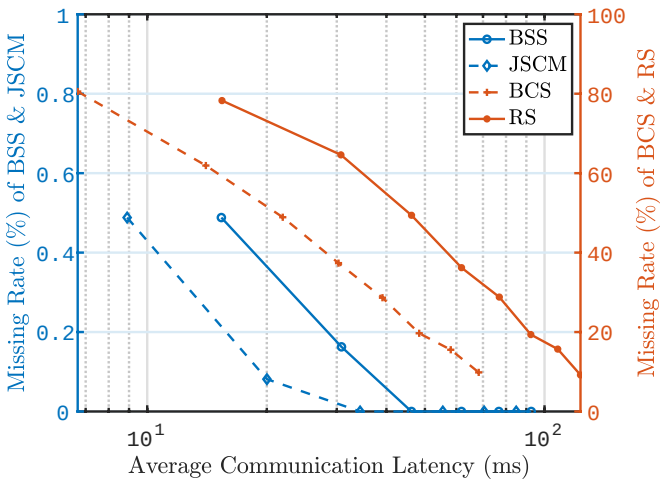


图5. 所提出的JSCM与三种基准方案——BSS、BCS和RS的丢失率随平均通信时延的变化。

### *B.* 性能评估

将 JSCM 的性能与 BSS、BCS 和 RS 进行比较。图 5 展示了丢失率随通信时延变化的曲线；随着选定传感器数量增加，通信时延也随之增长。我们可以看到，作为一种以速率为中心的方案，BCS 和 RS 都因缺乏语义感知而表现不佳，性能无法接受。另一方面，由于具备这种感知能力，JSCM 和 BSS 方案的丢失率远低于前述方案，例如，在 30 毫秒时延下，丢失率低两个数量级以上。在 JSCM 和 BSS 之间，所提出的设计显著优于后者。这表明 JSCM 是一种很有前景的解决方案，也说明了面向任务的无线设计的必要性。

## VI. 总结

我们提出了SEMDAS框架，以解决边缘智能应用场景中的数据采集所导致的6G系统通信瓶颈问题。其基本原理是仅传输语义与计算任务相匹配的数据，从而避免传输无关数据造成的冗余网络流量。基于这一原则，我们提出了一个综合框架，其中包括架构、协议、基于学习的语义匹配技术，以及面向任务的无线技术的新设计原则。从更高层面来看，SEMDAS为6G语义通信系统的持续发展作出了贡献。具体而言，该新框架指明了变革无线网络架构的新方向，以支持高效的边缘智能操作与应用。

## 参考文献

[1] 华为，《6G：从连接人与物到连接智能的下一个前沿》，[在线] https://www-file.huawei.com/-/media/corp2020/pdf/tech-insights/1/6g-white-paper-en.pdf?la=en（访问日期：2022年12月1日）。

[2]《经济学人》，《世界上最有价值的资源不再是石油，而是数据——监管互联网巨头》，[在线] https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data（访问日期：2022年12月1日）。[3] 爱立信，《爱立信移动报告》，[在线] https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2022（访问日期：2022年12月1日）。[4] G. Zhu、D. Liu、Y. Du、C. You、J. Zhang和K. Huang，《迈向智能边缘：无线通信与机器学习相遇》，电气电子工程师学会通信杂志，第58卷第1期，第19–25页，2020年1月。[5] W. Weaver，《对通信数学理论的近期贡献》，*ETC*：普通语义学评论，第10卷第4期，第261–281页，1953年。[6] M. Jankowski、D. G¨und¨uz和K. Mikolajczyk，《边缘无线图像检索》，电气电子工程师学会精选领域通信期刊，第39卷第1期，第89–100页，2021年11月。[7] Z. Qin、X. Tao、J. Lu、W. Tong和G. Y. Li，《语义通信：原理与挑战》，[在线] https://arxiv.org/pdf/2201.01389.pdf（访问日期：2022年12月1日）。[8] Q. Lan、D. Wen、Z. Zhang、Q. Zeng、X. Chen、P. Popovski和K. Huang，《什么是语义通信？机器智能时代传达含义的观点》，通信与信息网络杂志，第6卷第4期，第336–371页，2021年12月。[9] Y. C. Liu、J. Tian、N. Glaser和Z. Kira，《When2com：通过通信图分组实现多智能体感知》，载于电气电子工程师学会/计算机视觉基金会计算机视觉与模式识别会议论文集（*CVPR*），美国华盛顿州西雅图，2020年6月13日至19日。[10] D. Wang、X. Cui、X. Chen、Z. Zou、T. Shi、S. Salcudean、Z. Wang和R. Ward，《基于Transformer的多视角三维重建》，载于电气电子工程师学会/计算机视觉基金会国际计算机视觉会议论文集（*ICCV*），加拿大魁北克省蒙特利尔，2021年10月10日至17日。[11] H. Yuan、W. Morningstar、L. Ning和K. Singhal，《我们在联邦学习中所说的泛化是什么意思？》，载于国际学习表征会议论文集（*ICLR*），加拿大魁北克省蒙特利尔，2022年4月25日至29日。[12] K. Huang、H. Wu、Z. Liu和X. Qi，《通过原位模型下载，在6G移动网络中实现多功能边缘人工智能》，[在线] https://arxiv. org/pdf/2210.03555.pdf，2022年。[13] R. Aljundi、P. Chakravarty和T. Tuytelaars，《Expert Gate：基于专家网络的终身学习》，载于电气电子工程师学会计算机视觉与模式识别会议论文集（*CVPR*），美国夏威夷州檀香山，2017年7月21日至26日。[14] M. M. Amiri和D. G¨und¨uz，《无线边缘的机器学习：空中传输的分布式随机梯度下降》，电气电子工程师学会信号处理汇刊，第68卷第1期，第2155–2169页，2021年3月。[15] F. Herzog、X. Ji、T. Teepe、S. H¨ormann、J. Gilg和G. Rigoll，《用于行人重识别的轻量级多分支网络》，载于电气电子工程师学会国际图像处理会议论文集（*ICIP*），美国阿拉斯加州安克雷奇，2021年9月19日至22日。

**Kaibin Huang** [电气电子工程师学会会士]，现任香港大学电气与电子工程系教授，工作地点为香港。他的研究兴趣包括移动边缘计算、边缘人工智能和6G系统。

**Qiao Lan** 于2019年获得深圳南方科技大学（南方科技大学）工学学士学位，目前正在香港大学攻读博士学位。

**Zhiyan Liu** 于2021年获得北京清华大学工学学士学位，目前正在香港大学攻读博士学位。

**Lin Yang** 获得香港科技大学博士学位，目前是华为诺亚方舟实验室的高级研究员。