

面向时间序列预测任务的异构联邦语义通信

邢焕来^{1),2),3)} 赵博文¹⁾ 宋富洪⁴⁾ 徐乐西⁵⁾ 类先富⁶⁾ 冯 力^{1),2),3)}

¹⁾(西南交通大学计算机与人工智能学院 成都 611756)

²⁾(西南交通大学唐山研究院 河北 唐山 063000)

³⁾(可持续城市交通智能化教育部工程研究中心 成都 611756)

⁴⁾(贵州财经大学信息学院 贵阳 550025)

⁵⁾(中国联通研究院 北京 100048)

⁶⁾(西南交通大学信息科学与技术学院 成都 611756)

摘 要 时间序列数据广泛存在于工业、医疗、交通等诸多物联网相关领域,蕴含着用户行为模式、系统运行状态、未来发展趋势等重要信息。传统的数据处理方法通常将终端设备采集的时间序列数据完整传输至数据中心,并通过集中分析实现智能决策。然而,随着时序数据的不断增多,海量数据传输将为传统通信网络带来巨大带宽压力,难以满足智能制造、智慧交通等场景的低时延、高可靠性等需求。为此,本文构建面向时序预测任务的异构联邦语义通信系统,通过提取并传输任务相关语义信息,提高时序数据传输效率。针对时序数据语义信息难以提取的问题,本文提出一种面向资源受限设备的多头混合线性注意力语义编解码算法,实现对全局一局部时序依赖关系的轻量化建模。针对异构设备间计算能力和数据分布差异导致的知识稀缺问题,本文提出一种联邦嵌入(Federated Embedding, FedEmb)算法,通过在系统内引入通用知识,促进异构编码模型高效训练和联合推理,提高语义通信系统的泛化性和鲁棒性。为进一步提高语义通信系统任务性能,本文提出基于自监督学习的二阶段预训练方案,通过依次对时序语义编码器、信道编码器进行生成式预训练,并对整个系统进行联合训练,充分挖掘传输数据中的任务相关语义信息。通过在五个真实世界时间序列数据集上开展仿真实验,结果表明, FedEmb 在面向时序预测任务的语义通信系统中相比四种异构联邦学习算法具有更好的知识共享效率和更低的均方误差。此外,与六种前沿的通信系统相比,本文提出的联邦嵌入语义通信(Federated Embedding based Semantic Communication, SC-FedEmb)系统具有鲁棒且高效的时序数据传输能力和最高的预测准确度,能够有效满足异构物联网环境中的实时预测需求。

关键词 语义通信;时间序列预测;异构联邦学习;注意力机制;自监督学习

中图法分类号 TP18

DOI 号 10.11897/SP.J.1016.2025.02207

Time Series Forecasting Oriented Heterogeneous Federated Semantic Communication

XING Huan-Lai^{1),2),3)} ZHAO Bo-Wen¹⁾ SONG Fu-Hong⁴⁾

XU Le-Xi⁵⁾ LEI Xian-Fu⁶⁾ FENG Li^{1),2),3)}

¹⁾(School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756)

²⁾(Tangshan Institute, Southwest Jiaotong University, Tangshan, Hebei 063000)

³⁾(Engineering Research Center of Sustainable Urban Intelligent Transportation, Ministry of Education, Chengdu 611756)

⁴⁾(School of Information, Guizhou University of Finance and Economics, Guiyang 550025)

⁵⁾(Research Institute, China United Network Communications Corporation, Beijing 100048)

⁶⁾(School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756)

Abstract Time series data are widely prevalent in various Internet of Things (IoT) domains,

收稿日期:2024-12-24;在线发布日期:2025-07-04。本课题得到国家自然科学基金(U23A20273, 62271420)、河北省自然科学基金(F2022105027)、中央高校基本科研业务费资助。邢焕来(通信作者),博士,副教授,博士生导师,中国计算机学会(CCF)会员,主要研究领域为语义通信、网络功能虚拟化、软件定义网络、人工智能、进化计算等。E-mail: hxx@home.swjtu.edu.cn。赵博文,博士研究生,主要研究领域为时间序列分析、语义通信、深度强化学习。宋富洪,博士,硕士生导师,主要研究领域为移动边缘计算和多目标优化。徐乐西,博士,教授级高工,主要研究领域为大数据、网络分析和智能运营。类先富,博士,教授,博士生导师,主要研究领域为 B5G/6G 无线通信理论与技术和下一代 Wi-Fi 关键技术。冯 力,博士,研究员,博士生导师,主要研究领域为网络空间安全与人工智能。

such as industrial manufacturing, smart healthcare, and intelligent transportation. These data often contain critical information such as user behavior patterns, system operating status, and future development trends. Traditionally, the raw time series data collected by numerous terminal devices are entirely transmitted to data centers, where computation-intensive data analysis and intelligent decision-making tasks are performed. However, as the volume of time series data continues to increase at an unprecedented rate, the centralized data processing paradigm brings tremendous bandwidth pressure to conventional communication systems and fails to meet the low-latency and high-reliability requirements of time-sensitive applications such as smart manufacturing and intelligent transportation. To tackle these challenges, this paper proposes a novel heterogeneous federated semantic communication system designed for time series forecasting tasks. Unlike traditional transmission methods, the proposed system leverages semantic communication principles, aiming to extract and transmit task-relevant information. This strategy significantly reduces communication overhead and achieves accurate forecasting-oriented time series transmission. One of the key contributions of this work is the design of a Multi-head Mixture Linear Attention (MMA) mechanism, which functions as the core component of the semantic encoders and decoders. MMA is tailored for the efficient extraction of semantic information in time series data. It combines the advantages of window attention and low-rank attention, enabling the modeling of both global and local temporal dependencies in a lightweight manner suitable for deployment on resource-constrained terminal devices. Another major challenge in semantic communication for time series forecasting is the scarcity of knowledge among distributed devices. Variations in computational capabilities and data distributions hinder the analysis and extraction of task-relevant semantic information on each device. To overcome this, this paper introduces a Federated Embedding (FedEmb) algorithm for effective knowledge sharing among heterogeneous devices. Through introducing common knowledge in the semantic communication system, FedEmb enables efficient training of the semantic encoders across distributed devices and facilitates collaborative inference without requiring direct access to raw data. This approach enhances model generalization and robustness in heterogeneous environments. To further enhance the overall performance of the semantic communication system, a two-stage pre-training scheme based on self-supervised learning is proposed. By sequentially performing generative pre-training on semantic encoders and channel encoders, the system can capture rich task-relevant representations and learn robust semantic features tailored to the forecasting objective. After that, the entire communication system undergoes joint optimization to ensure end-to-end performance improvement and alignment between semantic encoding and downstream forecasting tasks. Extensive simulation experiments are conducted on five real-world time series datasets to evaluate the proposed system. The results show that FedEmb outperforms four state-of-the-art heterogeneous federated learning algorithms, achieving superior knowledge sharing efficiency and significantly lower mean squared error (MSE) in time series forecasting tasks. Furthermore, compared with six advanced semantic and traditional communication systems, the proposed FedEmb-based Semantic Communication (SC-FedEmb) system exhibits enhanced robustness, reduced transmission overhead, and the highest forecasting accuracy under various real-world conditions. These findings demonstrate the effectiveness of SC-FedEmb in addressing real-time forecasting needs in heterogeneous IoT environments.

Keywords semantic communication; time series forecasting; heterogeneous federated learning; attention mechanism; self-supervised learning

1 引言

随着物联网技术的普及,时间序列数据呈指数级增长。在此背景下,对时间序列进行实时分析与精准预测变得尤为重要。时间序列预测是指通过分析历史时间序列数据的趋势、周期性等内在模式,构建数学模型以推测未来时间点数值或状态的过程,在工业、医疗、交通等领域有着广阔的应用^[1],如图1所示。以智慧交通场景为例,时间序列预测可以对交通流量进行实时监测与预警。通过分析历史的交通流量数据,预测特定路段在未来的交通情况,以提前采取调度措施,减少交通拥堵,提高道路通行效率。



图1 时序预测应用场景

传统的时间序列预测方法通过终端设备采集时间序列数据,并将其传送到数据中心进行分析和处理。然而,由于时间序列数据实时产生,对海量时间序列数据进行不间断的传输会为传统通信系统带来巨大负担。实际上,由于时序数据存在明显的冗余性,对收集到的完整时间序列数据进行传输,通常是不必要的。例如,在工业异常检测场景下,单一工业设备并不会频繁发生异常状况。因此,对海量时间序列数据的完整传输会造成大量的通信资源浪费。

语义通信作为一种前沿的数据传输技术,通过分析原始数据与下游任务间的关联,提取并传输数据中任务相关的语义信息,能够极大地节省通信资源消耗,并高效完成接收端任务。目前针对语义通信的研究,通常根据其下游任务的不同,分为面向数据重建的语义通信和面向智能任务的语义通信。随着人工智能及物联网等智能应用的发展,面向智能任务的语义通信技术受到了研究者们广泛的关注。当前,面向时序预测任务的分布式语义通信技术尚处于起步阶段,缺乏相关研究,面临诸多挑战。

(1)时序特征提取难与时序编码复杂度高。现有语义编解码研究多面向计算机视觉、自然语言处

理等领域,缺乏对时间序列局部性、周期性、冗余性以及噪声干扰等特性的相关考虑,无法实现高效的时序特征挖掘和传输,从而导致时序数据语义通信性能受限。此外,当前具有高性能的注意力类算法通常具有较高的复杂度,难以部署在移动设备等资源受限终端,限制了语义通信技术的发展。

(2)分布式语义通信异构设备知识共享难。不同于集中训练和推理的传统时序预测场景,分布式语义通信常面临单一设备数据匮乏、多设备间难以合作推理等问题,导致系统性能受限,无法实现高效的时序预测。现有研究假定设备间编码模型具有相同架构,可通过联邦平均等传统联邦学习算法计算不同设备编码模型参数的均值,进而实现知识共享和联合推理。然而,在实际场景中,不同设备编码模型间可能存在显著差异,如神经网络的层数,单层网络的参数量等。多数联邦学习算法难以适配上述异构场景,导致模型间知识难以共享,严重降低了模型训练效率和推理性能。

(3)现有训练方法无法有效提高语义通信性能。现有语义通信的相关工作通常对语义编码器和信道编码器联合训练。然而,由于缺乏对时序预测任务的针对性语义编码训练,上述方案无法充分发挥深度学习的表征提取能力,难以从时序数据中提取面向预测任务的高质量语义信息。此外,相比于图像等数据,时序数据易受噪声干扰,时序数据的周期、趋势等信息可能因部分异常值而发生显著改变。因此,联合训练方案会限制信道编解码能力,改变接收端语义信息,进而影响系统最终性能。

为解决上述问题,本文提出一种面向边缘网络时序预测任务的多对一异构联邦语义通信系统。多个终端设备采集不同种类的时序数据,通过无线信道传输至边缘服务器进行联合推理。本文致力于提高终端设备时序编码性能,实现异构设备间知识共享,提高时序数据语义通信训练效率,最终提高接收端时序预测任务的准确率。本文的主要贡献总结如下:

(1)针对面向时序预测任务的海量数据传输问题,构建了面向时序预测的异构联邦语义通信系统。通过挖掘并传输不同设备中时序数据的任务相关语义信息,提高数据传输效率,实现异构设备间的联合推理。

(2)针对时序数据中复杂的特征依赖关系,以及传统注意力机制的编码效率问题,设计了面向语义通信的多头混合线性注意力(Semantic Communi-

cation-oriented Multi-head Mixture Linear Attention, SC-MMA) 时序编解码模型, 以分析时序数据的局部性、周期性和冗余性特征, 并提出基于“趋势—季节”分解的时间序列传输方案, 实现时序数据的鲁棒传输。

(3) 针对分布式系统中异构编码模型间知识共享难, 以及训练效率和推理性能低的问题, 设计了联邦嵌入(Federated Embedding, FedEmb)算法, 通过共享不同设备的联邦嵌入参数, 实现异构编码模型间的知识共享, 保障分布式语义通信系统的整体性能。

(4) 针对训练数据稀缺导致的模型性能受限问题, 基于自监督学习, 设计了面向时序数据语义通信的“时序预测—特征修复”二阶段预训练范式, 同时提高时序语义编码和信道编码的模型性能, 进而提高接收端时序预测任务的准确率。

(5) 通过对多种真实场景的时间序列数据进行实验, 结果表明, 本文提出的异构联邦语义通信系统(SC-FedEmb)超越了六种前沿的基准模型, 在接收端时序预测任务中具有较优性能。

本文接下来组织如下: 第2节概述了语义通信、时间序列预测和联邦学习的相关工作; 第3节与第4节分别介绍了系统模型和 SC-FedEmb 的原理与实现细节; 第5节给出了仿真结果与对比分析; 第6节对本文的研究内容进行了总结。

2 相关工作

对时间序列的高效传输涵盖了语义通信、时间序列预测、联邦学习等多个领域, 本节将从上述三个方面回顾相关工作的研究进展。

2.1 语义通信

根据传输数据种类的不同, 目前对于语义通信技术的研究可以分为三类: 面向计算机视觉、面向自然语言处理, 以及面向多模态等数据的语义通信。

2.1.1 面向计算机视觉

基于信息瓶颈理论, Shao 等人^[2]研究局部特征提取和面向下游任务的分布式特征编码, 消除边缘设备数据的冗余, 实现了低延迟联合推理。通过设计选择性重传机制, 进一步降低通信代价。Xie 等人^[3]基于分层视觉 Transformer 构建了联邦语义通信系统, 通过任务自适应翻译器实现物联网场景下的分布式多任务图像传输。Guo 等人^[4]研究了元宇宙中虚拟现实传输对传统通信网络的压力, 基于单

视图数据训练三维感知生成模型, 并构建联邦多视图合成架构, 以提高模型训练效率。为权衡传输过程中的语义信息和细节信息, Tang 等人^[5]设计了语义对比损失和语义对比编码技术, 并提出了基于对比学习的语义通信系统。通过引入语义重编码方案, 该系统实现了在后续任务模型未知情况下对编码模型的有效预训练。Dong 等人^[6]提出了基于分层的语义通信系统, 将语义编解码模型分为基础模型和增强模型。基础模型通过多尺度生成对抗模型进行训练, 增强模型用于适配不同情况下的语义传输任务。Zhang 等人^[7]设计了语义失真中断概率作为性能指标以衡量语义通信的可靠性问题, 并基于此提出了具有自适应信道反馈的语义通信框架, 通过主动调整通信资源开销提高图像传输可靠性。

2.1.2 面向自然语言处理

Xie 等人^[8]围绕物联网设备资源受限问题, 提出了基于剪枝和模型量化技术的轻量化分布式语义通信系统, 显著改善了物联网文本数据的传输效率。Hu 等人^[9]面向广播场景设计了一对多语义通信系统, 基于迁移学习和知识蒸馏, 加速新用户解码模型的训练速度。Nan 等人^[10]分析了分布式深度学习模型训练中的样本缺乏问题, 设计了 UDSem 架构, 以实现面向设备算力的学习机制和基于模块切分的联邦优化方案。为实现鲁棒的语义通信, Peng 等人^[11]提出了语义校正器以辅助语义编码, 并基于自适应生成器和非自回归架构设计了并行语义解码器以提高解码效率。

2.1.3 面向多模态数据

Xie 等人^[12]围绕视觉问答任务, 设计了分层 Transformer 算法, 逐层提取并融合文本和图像特征。为进一步提高语义编码的通用性, Zhang 等人^[13]提出了面向多模态数据的统一语义通信系统。通过设计向量级动态方案和轻量级特征选择模块, 实现不同信道环境下传输特征量自适应调整, 进而优化传输效率。Wang 等人^[14]提出了一种用于增强现实和虚拟现实的多模态分布式语义通信系统, 该系统采用了双向缓存模型, 通过将缓存过程建模为马尔可夫决策过程, 设计基于内容流行度的深度 Q 网络, 以最小化时延、能耗等目标。此外, 还有少数如文献[15-16]等面向语音传输和语音识别相关任务的分布式语义通信研究。

现有关于语义通信的研究多集中于图像和文本数据的语义信息提取与传输。然而, 随着人工智能、物联网等领域的飞速发展, 时间序列预测已成为越

来越多智能应用的支撑技术^[17],研究面向时序数据的语义通信同样意义重大。目前,海量时序数据的传输和分析正为传统通信技术带来前所未有的挑战,而现有面向时间序列数据的语义通信研究尚浅。

2.2 时间序列预测

时间序列预测技术在工业、交通、气象等领域有着广阔的应用前景。学术界关注的主流时间序列预测算法多基于注意力机制或传统深度学习技术。

2.2.1 基于注意力机制

Informer^[18]通过概率稀疏自注意力机制对时间序列的重要时间点进行全局建模,并提出生成式解码器,显著提高了时间序列的预测效率。PatchTST^[19]提出了通道独立性假设,并对不同变量的时间切片进行分别建模。Crossformer^[20]通过对时间切片合并,在编码器端生成具有不同尺寸的时间片特征,并通过注意力机制直接建模时间序列变量间的关系。iTransformer^[21]对标准 Transformer 编码方法进行革新,通过线性映射直接建模时序依赖关系,并通过注意力机制建模线性映射后的变量特征,实现了高效的时间序列预测。

2.2.2 基于传统深度学习技术

N-BEATS^[22]仅通过全连接网络和残差结构进行时序建模,实现了优于传统时序预测技术的性能表现。Autoformer^[23]利用深度分解架构逐层分解趋势与季节性特征,并基于快速傅立叶变换分析序列的自相关性。DLinear^[24]对时间序列进行分解,并对趋势与季节性特征分别进行预测以实现简单高效的时间序列预测。MICN^[25]基于多尺度等距卷积块实现对全局一局部特征的有效提取。

现有研究多基于单一设备或集中式处理对时间序列的时间维度和变量维度进行分析,难以适配资源受限场景下的数据持续传输需求。分布式场景下,传统编码器-解码器架构难以实现有效的语义特征传输,且上述方法多难以实现对分布式设备数据的联合建模与推理。因此,研究面向分布式系统的时间序列预测技术具有重要价值。

2.3 联邦学习

联邦学习通过在不同设备之间共享模型参数、软标签等信息,实现分布式系统中的知识共享与隐私保护,在现实生活中有着广泛的应用价值^[26]。根据不同设备模型间架构的差异,联邦学习可分为模型同构与模型异构两类。

2.3.1 模型同构联邦学习

作为联邦学习的早期研究,联邦平均^[27](Fed-

erated Averaging, FedAvg)算法是当前诸多前沿联邦学习研究的基础。FedAvg 平均所有上传至参数服务器的模型参数实现模型聚合,在不共享训练数据的条件下,提高分布式系统中模型的训练效率并实现知识共享。联邦近端算法^[28](FedProx)在 FedAvg 的基础上,通过近端项约束参数优化过程,缓解不同模型收敛速度差异带来的影响,实现分布式系统中模型的稳定收敛。随机控制平均算法^[29](Stochastic Controlled Averaging for Federated Learning, SCAFFOLD)引入控制变量调节不同模型局部更新方向,提高系统整体收敛速度。为缓解不同终端训练数据非独立同分布造成的特征漂移问题,联邦批归一化算法^[30](FedBN)通过在联邦聚合前进行局部批归一化,并在联邦学习过程中冻结批归一化层,以实现加速收敛的效果。联邦简易表征算法^[31](Federated Simple Representation, FedSR)通过基于 L2 范数和条件互信息的正则化技术,避免模型学习联邦模型间的无效关联,提高模型对新用户的泛化能力。

2.3.2 模型异构联邦学习

作为对异构联邦学习的早期探索,联邦蒸馏算法^[32](Federated Model Distillation, FedMD)通过平均不同模型对公开数据集样本输出的软标签,实现相同样本空间中,异构模型的特征对齐,以间接实现知识共享。不同于 FedMD,联邦互相关学习^[33](Federated Cross-Correlation Learning, FCCL)算法构建基于公开数据集样本的互相关矩阵,促进异构模型不同特征维度的多样性。FCCL 通过联邦自监督学习,对齐异构模型间的特征维度,实现高泛化性的联邦优化。联邦异构神经网络^[34](Federated Heterogeneous Neural Networks, FedHeNN)通过中心核对齐方法衡量异构神经网络表征间的距离,以约束异构模型更新过程。联邦表征学习^[35](Federated Representation Learning, FedRep)要求低层网络共享相同架构,通过参数平均实现表征共享,并基于高层异构网络实现不同设备的个性化表征。联邦滚动子模型提取算法^[36](Federated Learning with Rolling Sub-Model Extraction, FedRolex)提出了滚动子模型提取方案,通过均匀训练全局模型的不同子结构,借助分布式系统有效训练大规模模型。针对联邦学习中的分类问题,联邦原型学习^[37](Federated Prototype Learning, FedProto)算法通过聚合不同设备的抽象类原型,共享全局类原型,正则化本地模型的训练过程,最小化局部模型分类误差。

现有研究面向同构模型的联邦学习多基于联邦平均算法实现高效的参数聚合^[38],而面向异构模型的相关研究多需要基于公开数据集实现表征对齐或假定部分模型结构同构,这两种方法在实际场景中通常有诸多限制。将其直接应用于分布式语义通信系统可能会导致训练效率或模型性能的明显下降。因此,如何在面向时间序列预测的分布式语义通信系统中实现知识的高效共享有待深入研究。

3 系统模型

如图 2 所示,本文构建了多对一语义通信系统,用于多变量时间序列协作预测。系统由多个移动设备和一个边缘服务器组成。移动设备作为单天线发送端,边缘服务器作为多天线接收端。在每个移动设备中,语义编码器从时间序列中提取语义信息,并

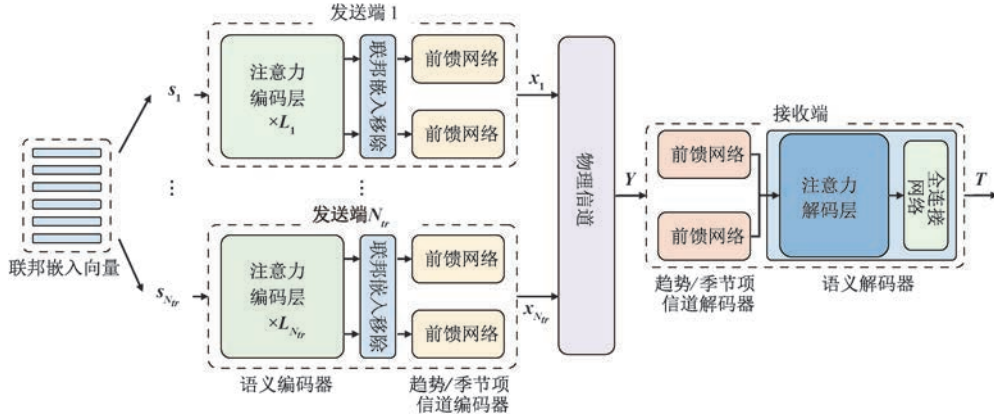


图 2 面向时序预测任务的分布式语义通信系统

通过信道编码器对语义信息进一步处理,以降低物理信道对语义信息的干扰。接收端接收到的信号通过信道解码器和语义解码器进行处理,以实现多变量时间序列协同预测。

3.1 发送端

图 2 为面向时序预测任务的分布式语义通信系统,假设系统包含 N_{tr} 个移动设备作为发送端,1 个边缘服务器作为接收端。每个移动设备收集不同种类的时间序列并挖掘对应的语义信息。记 $s_i \in \mathbb{R}^{L_{ts} \times N_v}$ 为移动设备采集的多变量时间序列,其中 L_{ts} 和 N_v 分别为 s_i 的长度和维度, $i=1,2,\dots,N_{tr}$ 。设 x_i 为设备 i 传输的语义特征,则语义编码与信道编码可定义为

$$x_i = C(S(s_i; A_i); \beta_i) \quad (1)$$

其中, $S(\cdot; A_i)$ 为用于提取时序信息的语义编码器, $C(\cdot; \beta_i)$ 为信道编码器。 $A_i = \{\alpha_i^l \mid l=1,2,\dots,L_i\}$ 和 β_i 分别为语义编码器与信道编码器的可训练参数。 α_i^l 表示编码器 i 第 l 层的网络参数, L_i 为编码器 i 网络层数,不同设备基于不同参数量的模型编码输入数据。设 A_g^t 和 A_i^{t+1} 分别表示服务器端全局参数与第 t 轮联邦学习后,设备 i 编码模型的参数,则联邦学习过程可定义为

$$A_g^t = \text{AGG}(A_1^t, A_2^t, \dots, A_{N_{tr}}^t) \quad (2)$$

$$A_i^{t+1} = \text{UPD}(A_g^t, A_i^t) \quad (3)$$

其中, AGG 为参数聚合函数,汇总所有客户端的知识。UPD 为本地参数更新函数,基于聚合的全局参数 A_g^t 更新本地模型。

3.2 物理信道

无线信道中,传输的语义信号会受到衰落和噪声的干扰。设 $Y \in \mathbb{C}^{N_{re} \times L_{si}}$ 为边缘服务器收到的信号,则无线传输可模拟为

$$Y = HX + Z \quad (4)$$

其中, $X \in \mathbb{C}^{N_{tr} \times L_{si}}$ 表示归一化后的复信号, L_{si} 表示信号向量的长度。 $H \in \mathbb{C}^{N_{re} \times N_{tr}}$ 与 $Z \in \mathbb{C}^{N_{re} \times L_{si}}$ 分别表示物理信道中的信道系数和加性高斯噪声。 N_{re} 为接收端天线数量。在加性高斯白噪声信道中, H 为单位矩阵。在衰落信道中, H 服从复高斯分布 $\mathcal{CN}(\mu, \sigma^2)$, μ 为均值, σ^2 为协方差。

3.3 接收端

接收端收集来自所有发送端的语义信号,并将整合后的全部信息依次馈入信道解码器与语义解码器。其中,信道解码器基于语义信息的上下文关联,缓解物理信道对语义信息的干扰,语义解码器分析还原后的语义信息,并在进一步时序特征提取后,完成系统的预测任务。记 $T \in \mathbb{R}^{L_{pr} \times (N_v \times N_{tr})}$ 为语义解码器最终预测的时间序列,其中 L_{pr} 为目标时间序

列长度,则信道、语义解码过程可表示为

$$\mathbf{T} = \mathbf{S}^{-1}(\mathbf{C}^{-1}(\mathbf{Y}; \omega); \phi) \quad (5)$$

其中, $\mathbf{C}^{-1}(\cdot; \omega)$ 和 $\mathbf{S}^{-1}(\cdot; \phi)$ 分别表示信道解码器和语义解码器, ω 和 ϕ 为可训练参数。

3.4 问题描述

前向传播后,系统通过最小化接收端预测损失 $Loss_{for}$ 优化各编、解码器的可训练参数。 $Loss_{for}$ 可衡量生成序列与目标序列之间的距离。

$$Loss_{for} = \text{MSE}(\hat{\mathbf{T}}, \mathbf{T}) \quad (6)$$

其中, $\hat{\mathbf{T}} \in \mathbb{R}^{L_{pr} \times (N_v \times N_{tr})}$ 表示预测的目标序列, MSE 为均方误差函数。系统优化中,接收端语义解码器与信道解码器通过常规的反向传播更新,发送端信道、语义编码器更新梯度可表示为:

$$\frac{\partial Loss_{for}}{\partial \beta_i} = \frac{\partial Loss_{for}}{\partial \mathbf{Z}} \frac{\partial \mathbf{Z}}{\partial \mathbf{Y}} \mathbf{H} \frac{\partial \mathbf{X}}{\partial \beta_i} \quad (7)$$

$$\frac{\partial Loss_{for}}{\partial A_i} = \frac{\partial Loss_{for}}{\partial \beta_i} \frac{\partial \beta_i}{\partial A_i} \quad (8)$$

其中, \mathbf{H} 为前向传播中随机生成的不可训练参数。基于学习率 η , 发送端 i 的参数 $\theta_i^{t_{lo}} \in \{A_i^{t_{lo}}, \beta_i^{t_{lo}}\}$ 在 t_{lo} 轮的更新过程为

$$\theta_i^{t_{lo}+1} = \theta_i^{t_{lo}} - \eta \nabla Loss_{for}(\theta_i^{t_{lo}}) \quad (9)$$

对于时序预测任务,语义信息通常包括关键周期和趋势性特征。例如,在能源领域,家庭用电量的日常变化情况可能反映出居民的生活方式与用电习惯,有助于构建更加精准的用户行为模型,实现对用电高峰时期能源消耗的准确预测。高效的语义编码器应能够过滤掉多种小功率设备短时运行导致的噪声波动,并准确传输高温或寒冷时期空调、供暖等设备的使用情况,从而为能源部门提供更具指导性的分析依据,助力科学决策与负载调度。传统卷积神经网络、循环神经网络类技术通常难以提取不同领域复杂时间序列数据的语义特征,导致数据传输效率和分布式模型训练效果的严重下降。因此,在分布式语义通信背景下探索多变量时序数据传输和预测方法至关重要。

4 面向时间序列预测任务的语义通信

本文设计了基于“趋势-季节”分解的轻量级多头混合注意力时序语义编码模型,并结合全连接网络(Fully Connected Network, FCN)进行语义解码。本文采用前馈神经网络(Feed-Forward Network, FFN)实现信道编解码以缓解物理信道干扰。

为高效更新不同客户端的异构模型,本文提出了一种基于联邦嵌入的异构联邦学习算法。为提高整个语义通信系统的训练效率,本文设计了面向时序语义通信的自监督预训练方案。

4.1 基于分解的轻量级多头混合注意力时序编码

为了实现面向预测的时序数据语义通信,发送端需要分析并提取任务相关语义特征,同时实现对信道干扰的鲁棒性。现有语义编码方案多面向图像、文本等数据,部分针对语音数据的相关研究,由于其与文本数据的高度相关性,噪声干扰,甚至部分数据的缺失,对整体语义影响有限。因此,现有少数相关研究多采用基础模型即可实现较好的性能。当前研究缺乏面向时序数据任务的针对性设计,忽略了时序数据的局部性、周期性、冗余性等特点。

来自不同发送端的时序数据,通常服从不同的概率分布,有着截然不同的周期性。例如,在智慧交通领域,一些路段车流量常年密集,一些路段常年车辆稀少。交通数据通常受随机性影响,不同时刻有着截然不同的特性。为充分挖掘时序数据普遍的局部性、周期性等特性,同时综合考虑移动设备可能存在的算力、存储能力的限制,本文设计了具有线性复杂度的多头混合注意力机制,并结合“趋势-季节”分解,实现面向时序数据的轻量化语义编解码模型,如图3所示。

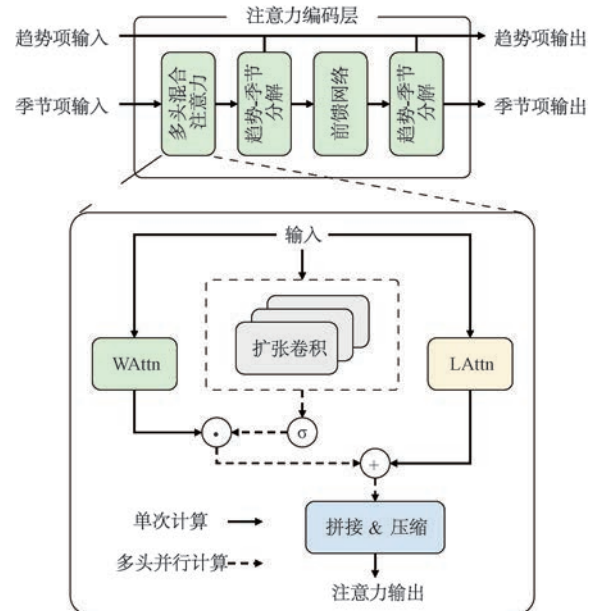


图3 轻量级多头混合注意力

4.1.1 窗口注意力

局部特征提取的质量有助于时间序列细节的预测,是影响预测准确性的重要因素之一。为了充分

挖掘时间序列中的局部信息,输入序列首先被切分为等长的时间片(Patch),并通过线性映射生成用于语义特征提取的嵌入序列。为了增强对时序数据局部信息的感知能力并探索短周期特征,本文对嵌入序列进行不重叠分段,段内通过窗口注意力(Window Attention, WAttn)进行语义提取,如图4所示。对于发送端*i*的语义提取,记 $p_{i,j}^{loc} \in \mathbb{R}^d$ 为窗口注意力聚合后,特征序列中位置*j*的局部特征, $j = 1, 2, \dots, L_{pt}$ 。 L_{pt} 为特征序列长度, d 为隐藏层维度,则窗口注意力公式表示为

$$p_{i,j}^{loc} = MSA(p_{i,j}^Q, p_{i,win(j)}^K, p_{i,win(j)}^V) \quad (10)$$

其中,MSA为多头注意力机制,以查询 $p_{i,j}^Q = W_Q^{loc} p_{i,j}$,键 $p_{i,win(j)}^K = W_K^{loc} p_{i,win(j)}$ 和值 $p_{i,win(j)}^V = W_V^{loc} p_{i,win(j)}$ 为输入。 W_Q^{loc} 、 W_K^{loc} 和 W_V^{loc} 为线性映射。 $p_{i,j} \in \mathbb{R}^d$ 为经过嵌入层映射后的输入特征, $p_{i,win(j)} \in \mathbb{R}^{k_{at} \times d}$ 为位置*j*所在窗口的子序列, k_{at} 为窗口尺寸。

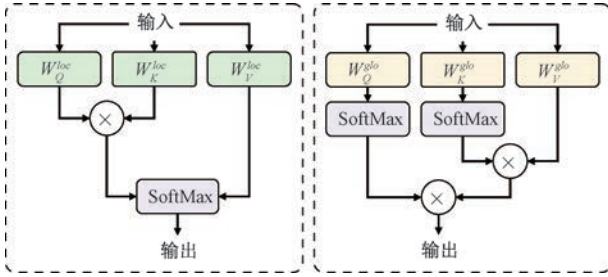


图4 窗口注意力(左)与全局注意力(右)

4.1.2 线性全局注意力

探索不同跨度周期间依赖关系以及时序数据的趋势变化,决定了需要提取的语义信息,并直接关系到接收端对语义信息重建效果,从而极大地影响接收端预测任务的性能。例如,周期为7与周期为17的特征是给定时间序列中的主要周期,则接收端对这两个周期的重建和预测与系统性能直接相关。传统注意力机制虽具有全局建模能力,然而其平方级的复杂度对资源受限的移动设备通常难以接受。本文基于对注意力矩阵的低秩性假设,引入线性复杂度注意力机制,如图4所示。记 $p_{i,1:L_{pt}}^{glo}$ 为聚合后的全局特征,则全局注意力的具体计算方法为

$$p_{i,1:L_{pt}}^{glo} = SM_h(\hat{p}_{i,1:L_{pt}}^Q) \mathbf{A}_i \quad (11)$$

$$\mathbf{A}_i = \frac{1}{L_{pt}} SM_v(\hat{p}_{i,1:L_{pt}}^K) \hat{p}_{i,1:L_{pt}}^V \quad (12)$$

其中, \mathbf{A}_i 为注意力矩阵, SM_v 与 SM_h 分别为对长度和维度进行归一化的SoftMax函数。 $\hat{p}_{i,1:L_{pt}}^Q$ 、 $\hat{p}_{i,1:L_{pt}}^K$ 与 $\hat{p}_{i,1:L_{pt}}^V$ 为经 W_Q^{glo} 、 W_K^{glo} 与 W_V^{glo} 映射后的特征序列。

4.1.3 多头混合注意力

假设注意力矩阵低秩虽然能显著降低复杂度,然而这通常会导致明显的性能损失。此外,传统多头注意力的计算量与参数量常随头数增多而成倍增长。本文通过多个扩张卷积,聚合不同范围的局部信息作为权重,并通过对窗口注意力与全局注意力的输出加权求和,在不显著提升参数量和计算量的同时,实现高效的多头混合注意力。记 $p_{i,1:L_{pt},h}^{mix}$ 为单次混合后的注意力输出,则

$$p_{i,1:L_{pt},h}^{mix} = w_h p_{i,1:L_{pt}}^{loc} + p_{i,1:L_{pt}}^{glo} \quad (13)$$

其中, h 为注意力头的索引, $h = 1, 2, \dots, N_h$, N_h 为注意力头的数量。权重 w_h 用于衡量不同位置局部特征对全局特征的重要性,可表示为

$$w_h = \sigma(DConv_h(p_{i,1:L_{pt}})) \quad (14)$$

其中, $DConv_h$ 为第*h*头对应的扩张卷积, σ 表示Sigmoid函数。多头混合注意力的输出 p_o 最后通过拼接concat和压缩映射 W_o 生成,表示为

$$p_o = concat(p_{i,1:L_{pt},1}^{mix}, p_{i,1:L_{pt},2}^{mix}, \dots, p_{i,1:L_{pt},N_h}^{mix}) W_o \quad (15)$$

4.1.4 “趋势—季节”分解

多数可预测时间序列可分解为趋势项、季节项。有效分析二者之间的关系,对于时序数据传输的鲁棒性和时序预测的准确性都是必要的。例如,在金融市场分析中,长期趋势与周期信息往往携带有价值的交易信号,而频繁的高频波动则更多被视为噪声。相反,在网络流量监控中,系统更关注秒级或分钟级的异常变化,长周期、趋势等往往会被忽略,因为对突发事件的快速检测和实时响应通常更加重要。

为解决上述问题,本文提出对时间序列进行“趋势—季节”分解的时序数据传输方案,如图3所示。在多头混合注意力以及前馈层后分别进行时序分解,将季节项作为后续网络的输入,将分解后的趋势项逐层累加。设 \bar{p}_i 为分解层的输入,则分解过程可表示为

$$\bar{p}_i^{tre} = AvgPool(\bar{p}_i) \quad (16)$$

$$\bar{p}_i^{sea} = \bar{p}_i - \bar{p}_i^{tre} \quad (17)$$

其中,AvgPool为平均池化函数。 \bar{p}_i^{tre} 与 \bar{p}_i^{sea} 分别为分解后的趋势项与季节项。通过两个不同参数的信道编码器对季节项和趋势项分别编码,以实现高效的抗干扰能力。

接收端采用两个基于前馈神经网络的信道解码器对趋势项和季节项分别解码,以实现精准的语义

特征重建,并将两项相加馈入语义解码器。由于接收端后续推理过程不需考虑信道干扰情况,语义解码器直接通过数层多头混合注意力与前馈网络对来自不同发送端的时序数据进行联合分析,最后通过单层全连接网络完成预测任务。

4.2 联邦嵌入算法

分布式语义通信系统中,终端设备通常受限于算力、存储、电量等资源,难以部署结构一致、参数量统一的编码模型。资源充足的设备通过部署高性能模型,可对输入数据进行复杂的长期依赖建模,资源受限的设备无法部署大规模深层模型,只能提取局部或浅层语义特征。设备间模型异构容易导致语义表示的不一致问题,且不同设备提取的语义特征对噪声的鲁棒性同样存在差异。此外,编码模型结构的差异导致系统难以通过传统联邦平均类算法在不同设备间实现知识共享,进一步限制了系统综合性能的提升空间。

现有面向异构模型的联邦学习主要基于公开数据集进行联邦蒸馏,以对齐本地模型的特征空间。但联邦蒸馏的效果直接受限于公开数据集的数据分布与数据质量。另一种常用的方法假设不同设备的模型具有部分相同的结构,从而可通过联邦平均等传统方法进行更新。这两种方法局限较多,难以应用于分布式语义通信场景,无法确保高效的分布式模型训练,并促进有效的协同推理。

为了促进异构编码模型之间的知识共享,本文设计联邦嵌入算法以实现语义通信系统跨设备知识共享,如图 5 所示。算法将映射后的联邦嵌入向量均匀插入编码模型嵌入层,联邦映射包含系统内的通用知识,可在语义编码的前向传播过程中提供丰富的上下文表示,实现分布式系统全局知识的高效共享。在反向传播中,联邦映射与联邦嵌入向量随模型参数一同更新,以从本地数据集学习重要语义知识。联邦嵌入算法的伪代码如算法 1 和算法 2 所示。

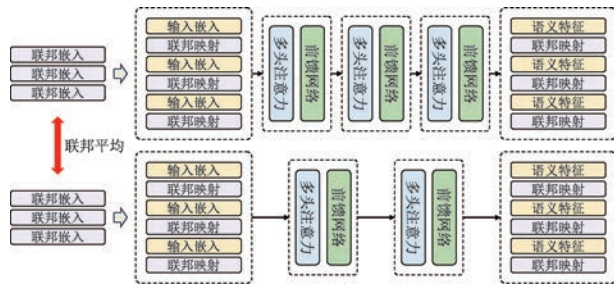


图 5 联邦嵌入算法示意图

算法 1. 联邦嵌入算法

输入: 输入序列 s_i 、发送端 i 的学习率 η_i 、联邦学习总轮次 T_{gl} 、本地更新总轮次 T_{lo} 、发送端 i 的联邦嵌入矩阵 F_i 、发送端 i 的参数 θ_i 、发送端 i 的本地数据集 \mathcal{B}_i 等算法参数。

输出: 发送端 i 的参数 $\theta_i^{T_{gl}}$

1. FOR 发送端 $i = 1, 2, \dots, N_{tr}$ DO
2. 初始化模型参数 θ_i^0
3. 根据公式(18)生成联邦嵌入参数 F_i
4. FOR 联邦学习轮次 $t_{gl} = 1, 2, \dots, T_{gl}$ DO
5. $F_i^{t_{gl}+1} \leftarrow$ 本地参数更新($\eta_i, T_{lo}, F_i^{t_{gl}}, \hat{F}_i^{t_{gl}}, \mathcal{B}_i$)
6. 根据公式(19)聚合各发送端联邦嵌入参数

算法 2. 发送端 i 本地参数更新

输入: 学习率 η_i 、本地更新总轮次 T_{lo} 、发送端 i 的联邦嵌入矩阵 F_i 、全局模型的联邦嵌入矩阵 \hat{F}_i 、发送端 i 的本地数据集 \mathcal{B}_i 等算法参数。

输出: 发送端 i 的联邦嵌入矩阵 F_i 。

1. 根据公式(20)对联邦嵌入进行动量更新。
2. FOR 本地更新轮次 $t_{lo} = 1, 2, \dots, T_{lo}$ DO
3. 采样一个批次的样本 $B_i \subset \mathcal{B}_i$
4. 从 B_i 中提取语义信息并进行信道编码
5. 经物理信道传输至接收端
6. 信道解码并进一步分析语义信息,完成预测
7. 根据公式(7)~(9)反向传播更新参数 $\theta_i^{t_{lo}}$ 与 $F_i^{t_{lo}}$

在系统模型初始化过程中,各客户端首先对联邦嵌入矩阵 $F_i \in \mathbb{R}^{d_{emb} \times L_{emb}}$ 进行线性映射,生成联邦映射矩阵 $E_i \in \mathbb{R}^{d_i \times L_{emb}}$ 。 d_{emb} 和 L_{emb} 分别为联邦嵌入的维度和长度, d_i 为客户端 i 的嵌入层维度,联邦映射矩阵的计算过程可表示为

$$E_i = \begin{cases} W_i^{fed} F_i, & \text{若 } d_i \neq d_{emb} \\ IF_i, & \text{若 } d_i = d_{emb} \end{cases} \quad (18)$$

若联邦嵌入矩阵 F_i 与客户端 i 的嵌入层具有不同维度,通过矩阵 $W_i^{fed} \in \mathbb{R}^{d_i \times d_{emb}}$ 对 F_i 进行映射,使之与客户端 i 的嵌入层维度一致。若不同客户端嵌入层维度相同,则可直接初始化 W_i^{fed} 为单位矩阵 I ,即 $E_i = F_i$ 。在将 E_i 均匀插入各客户端嵌入层后,联邦嵌入矩阵 F_i 可随语义通信的前向传播与反向传播更新,如图 6 和图 7 所示。

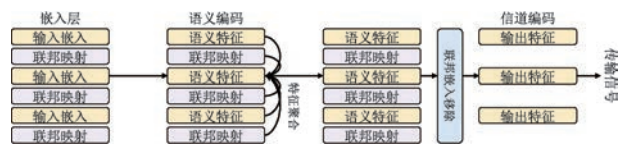


图 6 联邦嵌入算法前向传播过程

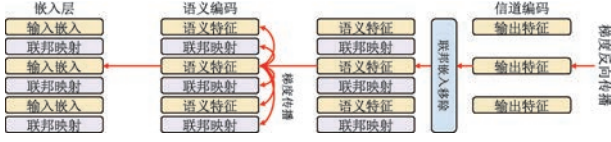


图 7 联邦嵌入算法反向传播过程

为了确保高效的协同推理,不同客户端的联邦嵌入矩阵 \mathbf{F}_i 之间采用参数平均方法共享知识,记 $\hat{\mathbf{F}}_i^{t_{gl}}$ 为轮次 t_{gl} 中聚合全局知识的联邦嵌入矩阵,其聚合过程如公式(19)所示。

$$\hat{\mathbf{F}}_i^{t_{gl}} = \frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} \mathbf{F}_i^{t_{gl}} \quad (19)$$

参数服务器在计算全局模型的联邦嵌入矩阵后,将参数下发至各发送端,各发送端基于动量 γ 更新本地联邦嵌入矩阵参数,如公式(20)所示。

$$\mathbf{F}_i = \gamma \times \mathbf{F}_i^{t_{gl}} + (1 - \gamma) \times \hat{\mathbf{F}}_i^{t_{gl}} \quad (20)$$

相比于现有异构联邦算法对高质量公开数据集的依赖或对部分网络同构的严格假设,联邦嵌入算法对网络结构限制小,参数更新效率高。联邦嵌入向量在设备间共享,能够直接在异构设备的时序语义特征提取过程中逐层传递系统任务相关的上下文信息,进而提高语义编码的性能。此外,联邦嵌入在语义编码后移除,信道编码器仅对其余输出特征进行处理,因此,联邦嵌入算法并不会显著增加语义通信的资源开销。

4.3 面向时序语义通信的自监督预训练

网络参数的初始化通常会直接影响模型后续更新的性能,现有语义通信训练方法多基于随机初始化参数对语义、信道编码器进行联合训练。然而,这种方法既无法充分挖掘时间序列中的各种特征,也会限制信道编解码器的特征重建能力。本文在语义、信道联合训练前,基于自监督学习,提出一种二阶段预训练方案,分别对语义编解码器和信道编解码器进行针对性预训练,以最大化面向时序预测任务的语义编码能力和面向语义特征的信道编码能力。

4.3.1 语义编码预训练

现有针对时序数据的预训练多采用类似掩码自编码器^[39]或BERT模型^[40]的掩码预训练方案。如在文献[19]中,研究者随机选择一部分时间片进行掩码后,基于注意力机制的双向编码能力进行生成,以使模型进一步了解时序数据的特性。然而该类方法忽略了时序预测任务的因果性,即在下游任务中,对一个时刻的数据预测,只能基于该时刻之前的信

息。因此,基于双向编码的生成式预训练并不适合时序预测任务。此外,语义通信以提取并传输时序语义信息为目标,因此还需要避免对一些时序数据中噪声波动的过度拟合。

为鼓励模型对长短周期、趋势等特征的提取,同时避免过度关注细小波动,本文设计面向时序数据的自监督语义编码预训练方案,如图8所示。在对输入嵌入添加高斯噪声之后,通过平均池化 $AvgPool$ 对特征进行平滑以去除微小波动,实现有效数据增强,丰富数据的周期、趋势特征,避免对时序噪声的过拟合。记 $p_{i,j}^{aug}$ 为平滑后的增强时序特征,则数据增强过程表示为

$$p_{i,j}^{aug} = AvgPool(p_{i,j} + \hat{Z}) \quad (21)$$

其中, \hat{Z} 服从 $\mathcal{N}(\mu_{aug}, \sigma_{aug}^2)$ 分布的高斯噪声。在对输入嵌入进行数据增强后,随机选择增强后的部分特征作为模型的输入,并对输入嵌入的部分后续位置进行掩码。在前向传播过程中,对整个序列所有位置的特征进行位置编码,并结合轻量级多头混合注意力机制建模局部与全局时序信息。语义编码预训练过程的优化目标为最小化生成损失 $Loss_{SE}$,即

$$Loss_{SE} = \sum_{i=1}^{N_{tr}} MSE(p_{i,mask}^{aug}, p_{i,mask}^{dec}) \quad (22)$$

其中, $p_{i,mask}^{aug}$ 为平滑后被掩码特征, $p_{i,mask}^{dec}$ 为接收端注意力解码层针对掩码位置生成的语义特征。在计算损失阶段,丢弃除掩码位置外其余全部特征,即只计算掩码位置对应特征的重建损失值。

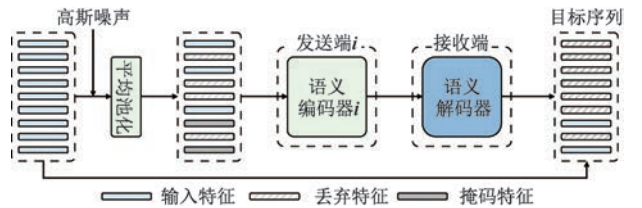


图 8 时序语义编码预训练示意图

语义编码预训练的前向传播过程中,每个发送端仅使用注意力编码层进行语义编码,并将其输出直接馈入注意力解码层实现语义解码。反向传播阶段根据注意力解码层对掩码位置的特征重建情况计算损失,并只计算注意力编解码层部分对应的梯度,以完成参数更新。语义编码预训练本质上是基于已知稀疏特征对后续掩码信息进行预测,通过增加生成任务的困难度,提升模型对上下文信息的理解能力,实现符合时序数据因果特性的自监督生成式预训练。

4.3.2 信道编码预训练

在对语义编解码器进行预训练后,需要根据训练好的语义编码空间进行针对性的信道编码,以最大化语义编码预训练后的特征表达能力。在冻结语义编码器参数后,将语义编码器提取的语义特征依次馈入信道编码器、物理信道以及信道解码器,进行去噪自编码训练,如图 9 所示。

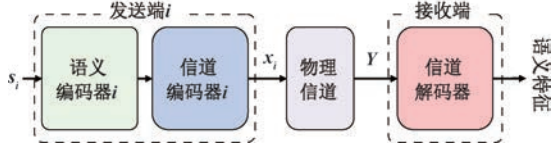


图 9 信道编码预训练示意图

信道编码预训练的优化目标为最小化语义编码器的输出 $p_{i,j}^{SE}$ 与信道解码器的输出 $p_{i,j}^{CH}$ 的距离,即

$$Loss_{CE} = \sum_{i=1}^{N_{tr}} MSE(p_{i,j}^{SE}, p_{i,j}^{CH}) \quad (23)$$

在整个预训练过程中,发送端语义编码器对输入时序数据进行语义特征提取和时序分解,而后对趋势项和季节项分别进行信道编码。接收端对收到的趋势项和季节项数据分别进行信道解码,并将二者相加实现语义信息重建,最终重建结果被馈入语义解码器进行相应的预训练任务。

在完成二阶段预训练后,系统在已训练好的语义编码器、信道编码器、信道解码器、语义解码器的基础上,继续进行语义—信道编码联合训练。语义解码器末端的全连接网络仅参与语义—信道编码联合训练,从而学习对目标序列的预测。

5 仿真结果与分析

本节首先介绍实验相关配置,包括数据集、参数设置和实现细节。然后,分别验证轻量级多头混合注意力时序编码算法和联邦嵌入算法的有效性。最后,在三种信道下对比本文提出的异构联邦语义通信系统与六种前沿通信系统的性能差异。

5.1 仿真设置

(1)数据集。为验证本文提出的语义编码模型、联邦学习模型和预训练策略的有效性,本文在五个广泛应用的真实世界时间序列数据集上进行实验,分别是 Exchange、ETT、Weather、Electricity 和 Traffic^[14]。Exchange 数据集记录了八个国家的汇率波动。ETT 数据集记录了变压器中负载、油温等指标随时间的变化数据。Weather 数据集记录 21 个气象设备每 10 分钟的指标变化,包含温度、气压、

风速等数据。Electricity 数据集包含数百名用户的每小时电力消耗记录。Traffic 数据集包含多个高速公路传感器每小时的数据变化,综合展示道路占用率变化趋势。

为模拟真实场景下的时间序列数据收集与传输情况,充分评估 SC-FedEmb 系统的综合性能,实验从每个数据集挑选了部分时间序列,并将这些序列均匀且随机地分配到各个设备上。为保证每个设备获得足够的训练和测试样本,不同种类数据传输实验配备不同数量的终端设备,如表 1 所示。为了评估传输质量和预测精度,本文假定语义通信场景为编码模型从长度为 336 的时间序列提取语义信息,接收端根据解码后的语义信息预测长度为 336 的未来时间序列,并通过均方误差直接衡量模型输出的预测误差。

表 1 实验数据配置

数据集	序列维度	设备数量
Exchange	8	2
ETT	7	7
Weather	21	7
Electricity	128	16
Traffic	128	32

(2)参数设置。本文采用加性高斯白噪声(Additive White Gaussian Noise, AWGN)、瑞利(Rayleigh)和莱斯(Rician)信道模拟无线环境,以测试传统通信系统和语义通信系统的性能。实验中信噪比(Signal-Noise Ratio, SNR)范围从 0 分贝(dB)到 20 分贝。对于 Rayleigh 信道,信道系数 H 服从复高斯分布 $\mathcal{CN}(0, I)$ 。莱斯因子 k_r 设置为 2。当训练环境与测试环境完全一致时,编码模型能发挥最佳性能。然而,受限于通信资源的稀缺性,为每一种信道环境分别训练对应的编码模型,并在通信过程中根据信道环境动态调整编解码方案是不现实的。为保证编码模型在不同信道条件下的稳定性,本文分别在加性白噪声信道、瑞利信道和莱斯信道下对模型进行训练,通过对比不同模型训练后在所有测试环境下的综合性能,选择综合性能最优的模型作为实际部署的编解码模型。

由于时序特征具有明确顺序关系,且邻近周期常具有一定相似性,因此本文采用旋转位置编码^[41]建模时序数据的相对位置信息,并在语义编码前和语义解码后使用实例归一化。在消融实验和综合性能评估中,设每个发送端嵌入层维度与多头混合注意力机制的隐藏层维度 d 默认均为 16,语义编码器中前馈神经网络的隐藏层维度默认为 64。信道编

码器中,输入特征维度与隐藏层特征维度分别为 16 和 32。联邦嵌入向量和联邦映射向量维度默认均为 64。联邦嵌入向量的数量默认与时间切片后特征向量的数量相同。在将原始序列采样为时间片的过程中,设时间切片长度为 8,采样步长为 8。注意力计算中,设窗口注意力的范围默认为 7。时序分解中滑动窗口长度为 25。每个发送端语义编码模型的网络层数从 1 到 4 随机选择。因不同种类时间序列的语义特征提取与预测难度可能存在较大差异,当传输数据语义信息较为复杂时,编码网络的隐藏层维度以及联邦嵌入向量的维度均需适当调高,以保证基本的语义通信能力。

(3)实现细节。神经网络基于 Pytorch 1.13 构建。实验中使用 Adam 优化器与均方误差训练所有语义通信模型。实验服务器配备 Ubuntu 20.04.2 操作系统、Intel Xeon(R) CPU E5-2667 v4 3.2 GHz 处理器、128GB 内存和配置 12G VRAM 的 Nvidia Titan V GPU。

5.2 轻量级多头混合注意力时序编码的有效性

为探寻轻量级多头混合注意力时序编码算法的有效性,本文将其与四种基于不同基础模型的语义编码算法进行对比,包括 DeepSC-S、DeepSC-ST、DL-SCMT 和 SC-PatchTST。因 SC-MMA 与 SC-PatchTST 为基于时间切片的算法,为消除不同特

征表示策略导致的性能差异,本文对 DeepSC-S、DeepSC-ST 和 DL-SCMT 进行与 SC-MMA 和 SC-PatchTST 相同的时间切片操作。五种编码算法采用完全相同的模型结构、学习率等超参数,并在 Weather、Electricity 和 ETT 三个数据集上进行测试。为评估不同编码算法的稳定性,测试信道选择干扰较为明显的莱斯信道。上述系统的具体实现细节如下所示:

(1)DeepSC-S:结合基于通道注意力的 SENet^[42]和残差网络进行语义编解码。

(2)DeepSC-ST:基于卷积神经网络和双向循环神经网络(Bi-GRU)进行语义编解码。

(3)DL-SCMT:采用残差网络进行语义编解码。

(4)SC-PatchTST:基于多头注意力机制进行语义编解码,并采用全连接网络进行时序预测。

(5)SC-MMA:本文所述基于分解的轻量级多头混合线性注意力语义通信编码算法。

如图 10 所示,DeepSC-S、DeepSC-ST、DL-SCMT 和 SC-PatchTST 算法在三种类型的时序数据传输场景下,性能表现均明显弱于 SC-MMA。本质上,在发送端对时序数据语义信息提取和在接收端对语义信息还原并执行下游任务的过程中,对时序数据的全局建模能够挖掘任务相关的季节性与趋势性信息,直接影响语义通信的性能与效率。

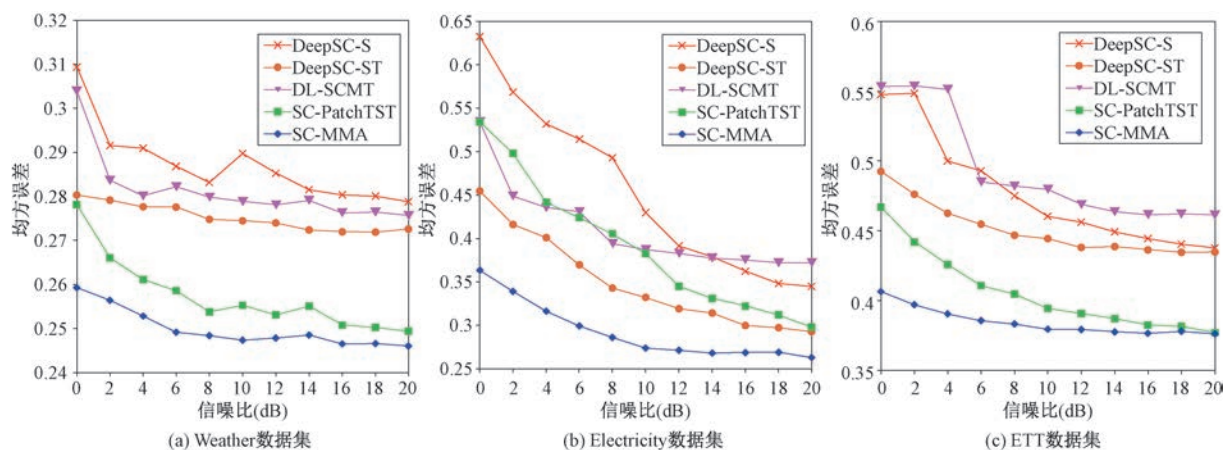


图 10 时间切片后五种系统的时序预测效果

在五中编码算法中,DeepSC-S、DL-SCMT 主要侧重局部信息编码,缺乏对数据全局信息的感知,性能相对较弱。DeepSC-ST 基于双向循环网络编码,因此具有一定的全局建模能力,但循环网络的天然缺陷导致其无法稳定对远距离依赖关系进行有效的分析。因此,其虽然在 Electricity 数据集上优于 SC-PatchTST,但在 Weather 和 ETT 数据集上均

明显弱于 SC-PatchTST。

基于注意力机制的 SC-PatchTST 能够直接建模全局信息,因此能够在 Weather 和 ETT 数据集下取得较好的性能表现。SC-MMA 结合低秩全局注意力、线性局部注意力能够更好地对任务相关的语义信息进行编码。此外,本文提出的基于“趋势—季节”分解的时序数据传输方案能够在接收端实现

更鲁棒的抗噪与语义信息还原效果。因此,SC-MMA能够实现优于四种基线系统的语义通信性能。

5.3 联邦嵌入算法的有效性

为了评估 FedEmb 算法的有效性,本文将其与四种广泛应用的异构模型联邦学习算法进行比较,包括 FedMD^[32]、FCCL^[33]、FedHeNN^[34]和 FedRep^[35]。实验在 ETT、Electricity 以及 Weather 三个数据集上进行,每个算法采用相同学习率并训练相同轮数。这五种算法的细节如下所示:

(1)FedMD:通过基于公开数据集的模型蒸馏对齐不同模型的输出特征,实现异构模型知识共享。

(2)FCCL:采用互相关矩阵,基于公开数据集对齐不同模型的特征维度,从而优化本地模型。

(3)FedHeNN:使用基于中心核对齐的近端项对模型参数正则化,从而提升语义编码器的性能。

(4)FedRep:假定所有客户端模型部分网络架构与参数一致,并通过联邦平均聚合参数。

(5)FedEmb:本文所述联邦嵌入算法。

表 2 展示了 ETT 数据集下五种联邦学习算法在时序数据语义通信系统中的效果。对比其他算法,FedEmb 在三种信道不同信噪比的实验中都实现了最优性能,得益于其参数更新效率,以及在前向传播中对整个网络的知识共享能力。在 AWGN 信道下,FedMD 略优于 FCCL 与 FedHeNN。在 Rayleigh 和 Rician 信道下,三种算法损失接近。这三种算法本质上都可视为基于公开数据集的特征对齐方法,因此具有相似的性能。除 Rayleigh 与 Rician 信道下的低信噪比情况,FedRep 预测损失最大,由于其更新参数无法影响整个网络,知识共享能力相对较弱。

表 2 五种联邦学习算法在 ETT 数据集均方误差(粗体表示损失最低,下同)

信道	算法	SNR=0	2	4	6	8	10	12	14	16	18	20	平均值
加性 白噪声	FedEmb	0.3772	0.3759	0.3750	0.3747	0.3743	0.3742	0.3741	0.3740	0.3740	0.3739	0.3739	0.3747
	FedMD	0.3809	0.3794	0.3789	0.3784	0.3781	0.3780	0.3778	0.3779	0.3778	0.3778	0.3778	0.3784
	FCCL	0.3811	0.3797	0.3791	0.3786	0.3783	0.3782	0.3781	0.3781	0.3780	0.3780	0.3780	0.3787
	FedHeNN	0.3821	0.3807	0.3798	0.3792	0.3788	0.3786	0.3784	0.3784	0.3783	0.3783	0.3782	0.3792
	FedRep	0.3842	0.3830	0.3823	0.3820	0.3818	0.3815	0.3814	0.3814	0.3814	0.3813	0.3813	0.3820
	FedEmb	0.3999	0.3904	0.3869	0.3842	0.3809	0.3767	0.3761	0.3760	0.3750	0.3745	0.3750	0.3814
瑞利	FedMD	0.4043	0.3990	0.3895	0.3860	0.3845	0.3820	0.3804	0.3791	0.3788	0.3790	0.3780	0.3855
	FCCL	0.4044	0.3992	0.3897	0.3862	0.3847	0.3809	0.3806	0.3793	0.3790	0.3792	0.3783	0.3856
	FedHeNN	0.4035	0.3961	0.3933	0.3872	0.3855	0.3820	0.3803	0.3802	0.3788	0.3788	0.3792	0.3859
	FedRep	0.4044	0.3980	0.3937	0.3917	0.3877	0.3844	0.3830	0.3834	0.3825	0.3824	0.3817	0.3884
	FedEmb	0.4011	0.3918	0.3873	0.3835	0.3814	0.3780	0.3770	0.3756	0.3753	0.3751	0.3738	0.3818
莱斯	FedMD	0.4080	0.3982	0.3905	0.3873	0.3841	0.3834	0.3813	0.3794	0.3793	0.3787	0.3785	0.3862
	FCCL	0.4082	0.3984	0.3907	0.3875	0.3843	0.3836	0.3815	0.3796	0.3795	0.3789	0.3787	0.3865
	FedHeNN	0.4090	0.3984	0.3938	0.3875	0.3855	0.3824	0.3809	0.3804	0.3804	0.3797	0.3788	0.3870
	FedRep	0.4054	0.3990	0.3939	0.3912	0.3880	0.3859	0.3852	0.3832	0.3826	0.3824	0.3824	0.3890

表 3 展示了五种算法在 Electricity 数据集中的性能差异。在 AWGN 信道中,FedEmb 具有显著低于其他联邦学习算法的预测损失。FedMD 与 FedHeNN 的性能位列第二和第三。FedRep 与 FCCL 算法性能相对较低。在干扰较强的 Rayleigh 和 Rician 中,FedEmb 在不同信噪比条件下仍具有最优性能。FedHeNN、FCCL 与 FedMD 效果接近。

表 4 展示了五种算法在 Weather 数据集的比较结果。FedEmb 在三种信道不同的信噪比下均获得了明显优于其他算法的最低预测损失。

综上所述,基于 FedEmb 的知识共享使得编解码模型对时序数据的预测准确度优于四个基线算法,证实了 FedEmb 在面向时序预测任务的异构语义通信中具有高效的知识共享能力。FedMD 等联邦蒸馏算法对公开数据集的质量依赖度较高,而 FedRep 对网络结构有一定限制,相比之下,FedEmb 具有较高的灵活性和应用价值。

dRep 对网络结构有一定限制,相比之下,FedEmb 具有较高的灵活性和应用价值。

5.4 整体性能评估

为了评估 SC-FedEmb 的有效性,本文将其与两种基于传统通信技术和四种基于语义通信技术的基线系统进行比较,包括 Trad-PTST、Trad-DLinear、FedSC-S、DisSC-ST、Dis-SCMT 和 Fed-PTST。实验在 Exchange、Weather、ETT、Electricity、Traffic 等五个数据集上进行,每个算法采用相同学习率并训练相同轮数。上述系统的实现细节如下所示:

(1)Trad-PTST:基于 Turbo 和 64-QAM 进行信道编解码和调制解调,并在接收端部署 PatchTST 模型进行时序预测。

(2)Trad-DLinear:基于 Turbo 和 64-QAM 进行信道编解码和调制解调,并在接收端部署 DLinear

表 3 五种联邦学习算法在 Electricity 数据集均方误差

信道	算法	SNR=0	2	4	6	8	10	12	14	16	18	20	平均值
加性 白噪声	FedEmb	0.2617	0.2596	0.2582	0.2574	0.2568	0.2565	0.2563	0.2561	0.2561	0.2560	0.2560	0.2573
	FedMD	0.2633	0.2612	0.2599	0.2592	0.2586	0.2583	0.2581	0.2580	0.2580	0.2579	0.2579	0.2591
	FCCL	0.2674	0.2656	0.2643	0.2637	0.2632	0.2629	0.2628	0.2627	0.2626	0.2626	0.2625	0.2637
	FedHeNN	0.2661	0.2642	0.2630	0.2623	0.2617	0.2615	0.2613	0.2612	0.2611	0.2611	0.2610	0.2622
	FedRep	0.2673	0.2654	0.2641	0.2635	0.2630	0.2627	0.2625	0.2624	0.2623	0.2623	0.2623	0.2634
	FedEmb	0.3454	0.3239	0.3015	0.2910	0.2825	0.2763	0.2707	0.2644	0.2625	0.2612	0.2608	0.2855
瑞利	FedMD	0.3522	0.3347	0.3175	0.3031	0.2936	0.2847	0.2753	0.2696	0.2687	0.2675	0.2647	0.2938
	FCCL	0.3572	0.3281	0.3146	0.3039	0.2958	0.2813	0.2776	0.2730	0.2677	0.2655	0.2659	0.2937
	FedHeNN	0.3566	0.3338	0.3174	0.3068	0.2945	0.2878	0.2757	0.2733	0.2690	0.2654	0.2645	0.2950
	FedRep	0.3570	0.3415	0.3278	0.3110	0.3038	0.2896	0.2848	0.2832	0.2767	0.2745	0.2728	0.3021
莱斯	FedEmb	0.3351	0.3116	0.3025	0.2892	0.2776	0.2728	0.2655	0.2639	0.2609	0.2590	0.2583	0.2815
	FedMD	0.3531	0.3214	0.3056	0.2987	0.2863	0.2836	0.2778	0.2690	0.2672	0.2668	0.2649	0.2904
	FCCL	0.3525	0.3214	0.3063	0.2999	0.2866	0.2842	0.2752	0.2747	0.2676	0.2664	0.2666	0.2910
	FedHeNN	0.3463	0.3231	0.3105	0.2992	0.2902	0.2798	0.2756	0.2668	0.2669	0.2650	0.2666	0.2900
	FedRep	0.3482	0.3218	0.3197	0.3022	0.2977	0.2847	0.2807	0.2786	0.2762	0.2727	0.2734	0.2960

表 4 五种联邦学习算法在 Weather 数据集均方误差

信道	算法	SNR=0	2	4	6	8	10	12	14	16	18	20	平均值
加性 白噪声	FedEmb	0.2461	0.2458	0.2456	0.2455	0.2454	0.2453	0.2453	0.2453	0.2453	0.2453	0.2453	0.2455
	FedMD	0.2506	0.2502	0.2500	0.2499	0.2498	0.2497	0.2497	0.2496	0.2496	0.2496	0.2496	0.2498
	FCCL	0.2501	0.2497	0.2496	0.2494	0.2493	0.2492	0.2492	0.2492	0.2492	0.2491	0.2491	0.2494
	FedHeNN	0.2492	0.2489	0.2487	0.2485	0.2485	0.2484	0.2484	0.2484	0.2483	0.2483	0.2483	0.2485
	FedRep	0.2498	0.2493	0.2491	0.2489	0.2488	0.2488	0.2487	0.2487	0.2487	0.2487	0.2487	0.2489
	FedEmb	0.2493	0.2505	0.2492	0.2474	0.2462	0.2463	0.2451	0.2450	0.2453	0.2452	0.2453	0.2468
瑞利	FedMD	0.2556	0.2523	0.2550	0.2552	0.2513	0.2498	0.2504	0.2506	0.2508	0.2500	0.2496	0.2519
	FCCL	0.2551	0.2518	0.2545	0.2547	0.2508	0.2493	0.2499	0.2501	0.2503	0.2496	0.2491	0.2514
	FedHeNN	0.2539	0.2517	0.2505	0.2494	0.2485	0.2496	0.2492	0.2488	0.2489	0.2485	0.2484	0.2498
	FedRep	0.2551	0.2535	0.2516	0.2487	0.2499	0.2491	0.2489	0.2491	0.2485	0.2495	0.2488	0.2502
莱斯	FedEmb	0.2498	0.2475	0.2477	0.2484	0.2469	0.2462	0.2456	0.2462	0.2454	0.2454	0.2452	0.2468
	FedMD	0.2600	0.2568	0.2563	0.2527	0.2522	0.2516	0.2502	0.2503	0.2498	0.2500	0.2499	0.2527
	FCCL	0.2595	0.2563	0.2559	0.2521	0.2516	0.2512	0.2497	0.2498	0.2493	0.2495	0.2493	0.2522
	FedHeNN	0.2582	0.2563	0.2515	0.2498	0.2497	0.2494	0.2491	0.2486	0.2491	0.2490	0.2490	0.2509
	FedRep	0.2549	0.2520	0.2503	0.2496	0.2507	0.2498	0.2501	0.2499	0.2492	0.2485	0.2485	0.2503

模型进行时序预测。

(3) FedSC-S: 基于 DeepSC-S 进行语义编解码, 并基于 FedHeNN 算法进行联邦学习。

(4) DisSC-ST: 直接微调 DeepSC-ST 实现分布式语义通信, 不通过联邦学习进行知识共享。

(5) Dis-SCMT: 直接微调 DL-SCMT 实现分布式语义通信, 不通过联邦学习进行知识共享。

(6) Fed-PTST: 基于 SC-PatchTST 进行语义编解码, 并通过 FCCL 进行联邦学习。

(7) SC-FedEmb: 本文所述基于分解的多头混合注意力和联邦嵌入算法的异构语义通信系统。

(8) SC-PFedEmb: 经过自监督预训练的 SC-FedEmb 算法。

图 11 展示了 Exchange 数据集上八种对比算法的均方误差。在 AWGN 信道下, SC-PFedEmb 与 SC-FedEmb 在不同低信噪比情况下始终具有最

优的性能表现。在信噪比大于 12 分贝的实验中, Fed-PTST 与上述两种算法性能接近。DisSC-ST、Dis-SCMT 与 FedSC-S 差异明显, 依次表现了 Bi-GRU、残差网络以及通道注意力在缺乏训练数据时对时序数据语义编码的性能。由于 Exchange 数据集规模较小, 模型训练不够充分, 鲁棒性相对较差, 导致语义通信系统在 Rician 信道下训练后, 在 AWGN 信道下的综合性能弱于 Rayleigh 和 Rician 信道。传统方法 Trad-PTST 和 Trad-DLinear 预测性能最弱。由于金融类数据的方差较小, 不同数据点间差异不明显, 轻微噪声都可能改变时间序列的波形, 传统编解码技术性能受限。

图 12 展示了八种语义通信系统在 Weather 数据集上的时序预测效果。在三种信道下, SC-PFedEmb 与 SC-FedEmb 取得了低于其他算法的均方误差。这得益于多头混合注意力对局部和全局语义特征的

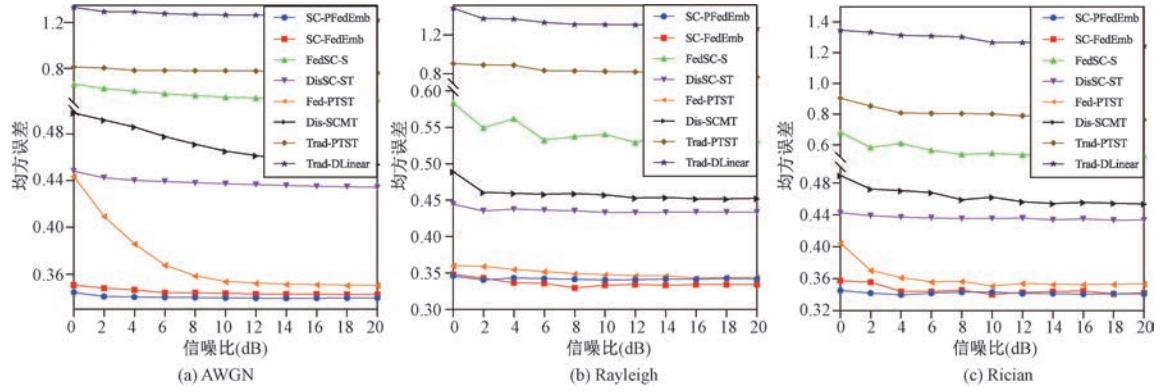


图 11 八种系统在 Exchange 数据集的时序预测效果

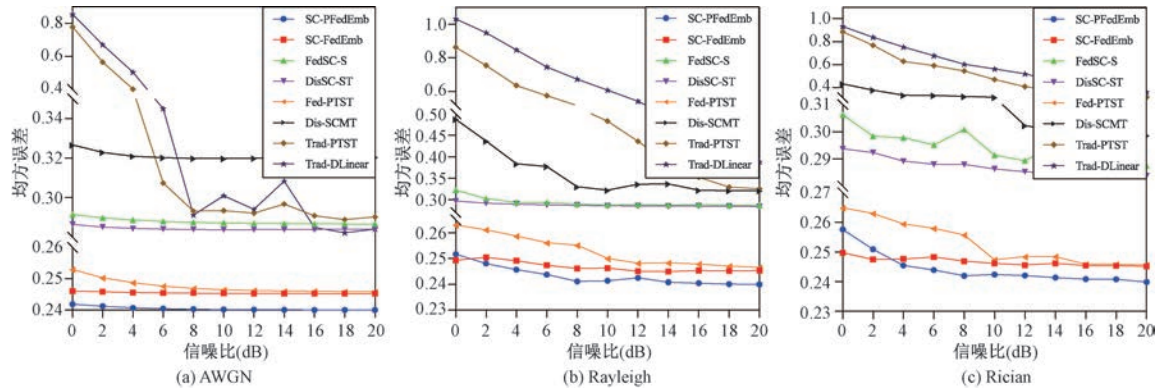


图 12 八种系统在 Weather 数据集的时序预测效果

有效建模,和时序分解在干扰下的鲁棒性。此外,SC-PFedEmb 在多数实验中实现了明显优于 SC-FedEmb 的性能表现,证实了自监督预训练的有效性。Fed-PTST 取得了仅次于上述两种算法的性能,且在高信噪比下实现了与 SC-FedEmb 相近的性能表现,证实了基于全局感知的注意力机制在语义通信中的优秀性能。DisSC-ST 与 FedSC-S 原用于语音处理系统,具有一定的时序编码能力,性能低于 Fed-PTST。Trad-PTST 与 Trad-DLinear 仅在高信噪比条件下能够实现相对可接受的预测效果。

图 13 展示了八种语义通信系统在 ETT 数据集

上的时序预测效果。由于 AWGN 下干扰较弱,Fed-PTST 在信噪比高于 12 分贝时,与 SC-FedEmb 的性能接近。而在 Rayleigh 与 Rician 信道下,Fed-PTST 系统仅在信噪比接近 20 分贝时与 SC-FedEmb 接近。此外,在三种信道的不同信噪比下,DisSC-ST 性能相对稳定。FedSC-S 在低信噪比下性能低于 DisSC-ST,仅在高信噪比下具有相对较好的性能表现。Dis-SCMT 因时序建模能力有限,在所有实验中性能最弱,在多数实验中弱于两种基于传统技术的基线算法,这也证实了研究面向时序数据传输的语义通信系统的必要性。

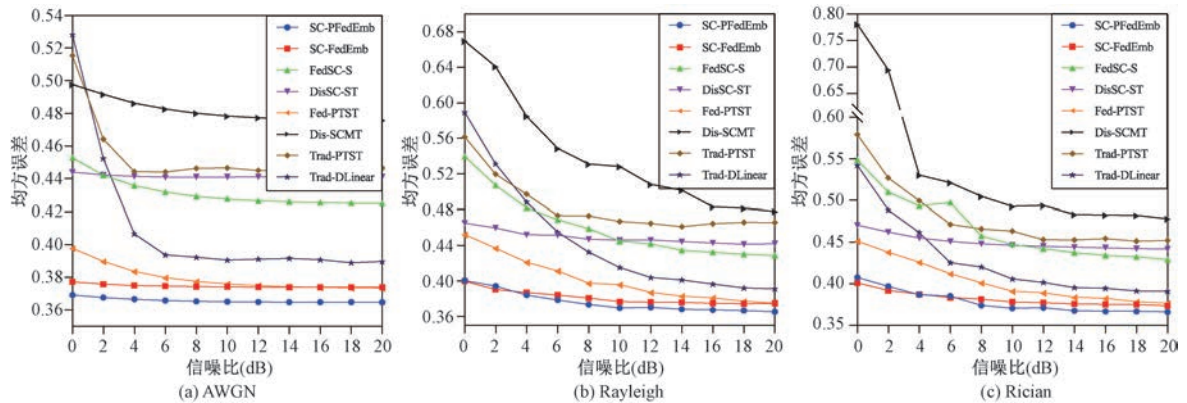


图 13 八种系统在 ETT 数据集的时序预测效果

图 14 展示了八种语义通信系统在 Electricity 数据集上的时序预测效果。SC-PFedEmb 与 SC-FedEmb 在所有实验中均展现了明显优势,在多数实验中 SC-PFedEmb 性能均优于 SC-FedEmb,展示了自监督预训练对语义编码和信道编码能力的提

升。上述两种系统效果显著优于其他所有基线系统。这充分展示了基于时序分解的语义编码对语义通信的显著作用。FedSC-S、DisSC-ST 以及 Dis-SCMT 性能相对较弱。Trad-DLinear 在信噪比较高的环境下具有与 Fed-PTST 相近的性能表现。

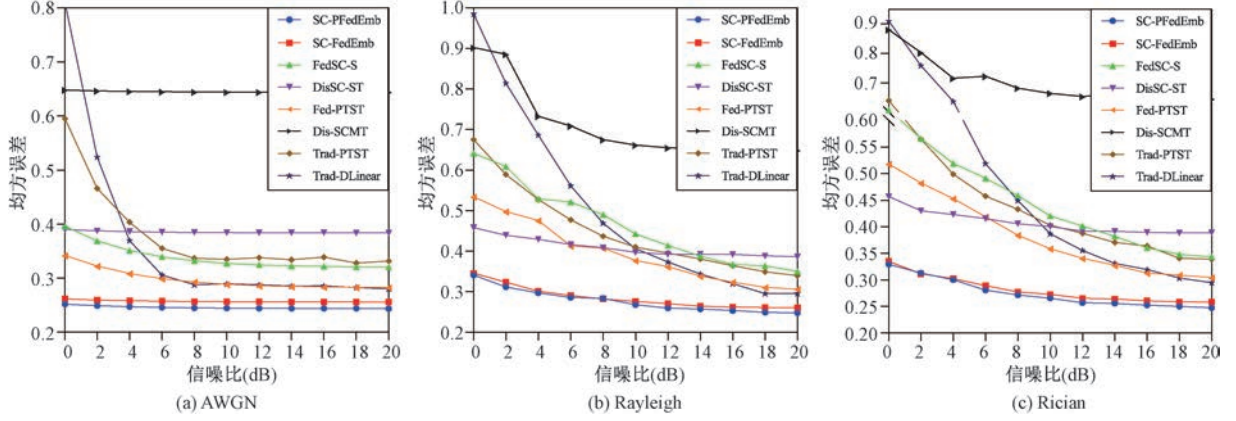


图 14 八种系统在 Electricity 数据集的时序预测效果

图 15 展示了八种语义通信系统在 Traffic 数据集上的时序预测效果。SC-PFedEmb 与 SC-FedEmb 在 AWGN 信道下性能明显优于其他四种语义通信系统。在 Rayleigh 和 Rician 信道中,DisSC-ST 和 Dis-SCMT 与上述两种算法在 0 分贝至 4 分贝的信噪比条件下,性能接近。随着信噪比增加,SC-PFedEmb 与 SC-FedEmb 性能优势愈加明显。在全部实验中,SC-PFedEmb 性能均优于 SC-

FedEmb,再次验证了自监督预训练的有效性。由于 Traffic 数据集规模较大,训练样本充分,Trad-PTST 与 Trad-DLinear 在信噪比较高的情况下能实现不弱于 FedSC-S 和 Fed-PTST 的性能表现。在强干扰情况下,源自前沿时序预测算法的 Fed-PTST 性能表现较弱,证实了现有时序预测算法无法直接应用于语义通信系统,难以发挥相对稳定的性能。

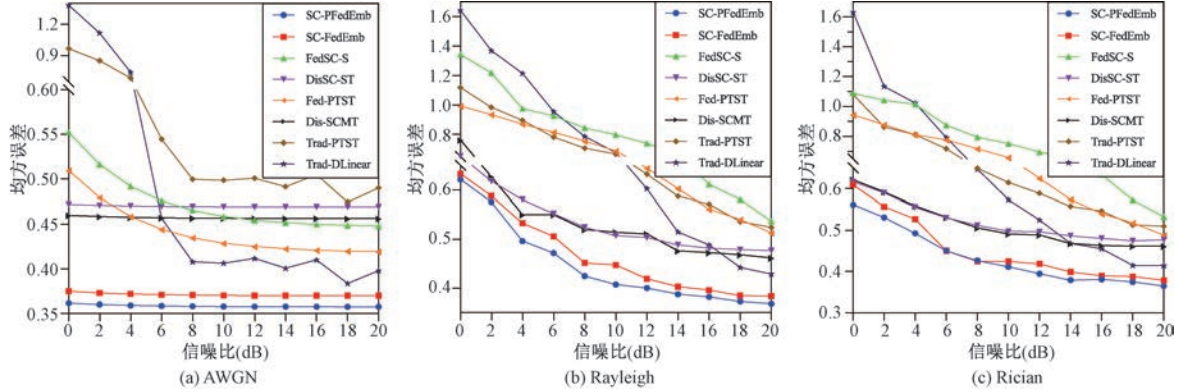


图 15 八种系统在 Traffic 数据集的时序预测效果

综上,在多数实验中,SC-PFedEmb 与 SC-FedEmb 具有最优的性能表现。Trad-PTST 与 Trad-DLinear 受数据集质量与规模影响明显,当样本充足、信道影响较弱时, PatchTST 与 DLinear 基于强大时序表征能力,展示了不弱于部分语义通信系统的性能表现。此外,当语义编解码器时序建模能力较弱时,即使采用语义通信的传输技术,系统仍难以实现面向智能任务的稳定数据传输,这证实了研究面向

时序数据语义通信系统的必要性。

5.5 语义编码复杂度分析

记时间序列长度为 L_{ts} , 时间片长度为 k_{pt} , 则嵌入层序列长度为 $L_{pt} = \frac{L_{ts}}{k_{pt}}$ 。记隐藏层维度为 d , 窗口长度为 k_{at} , 则窗口注意力复杂度为 $O(N_v \times \frac{L_{pt}}{k_{at}} \times (k_{at} \times d^2 + k_{at}^2 \times d))$, 全局注意力复杂度为

$O(N_v \times L_{pt} \times d^2)$ 。记多头混合注意力的头数为 N_h , 扩张卷积核的长度为 k_{di} , 则多头混合注意力整体的复杂度为 $O(N_v \times L_{pt} \times (k_{at} \times d + k_{di} \times d^2 \times N_h))$ 。记“趋势-季节”分解中采用长度为 k_{de} 的池化核, 则分解操作的复杂度为 $O(N_v \times L_{pt} \times d \times k_{de})$, 由于 $N_v, N_h, k_{pt}, k_{at}, k_{di}, k_{de}$ 以及 d 不会实时增长, 通常可视为常数。因此, 语义编码的整体复杂度可表示为 $O(L_{ts})$ 。

6 结 论

本文提出了一种面向时间序列预测任务的联邦语义通信系统 SC-FedEmb, 用于分布式设备间的协同推理。本文设计的基于分解的轻量级多头混合注意力机制能够识别不同尺度趋势和周期的重要性, 并具有对信道干扰的鲁棒性。本文提出的联邦嵌入算法显著提高了异构语义通信系统发送端的知识共享效率, 并实现了高性能的协同推理。本文提出的面向时序语义通信的自监督预训练方法, 显著提高了整个语义通信系统的训练效率和预测准确性。实验结果表明, 联邦嵌入算法优于四种前沿的异构联邦学习算法。此外, SC-FedEmb 在面向预测任务的语义通信中, 对比六种前沿的基线方法, 具有最优的时间序列数据传输性能。

参 考 文 献

- [1] Li Z X, Liu H Y. Combining global and sequential patterns for multivariate time series forecasting. *Chinese Journal of Computers*, 2023, 46(1): 70-84. (in Chinese)
(李兆玺, 刘红岩. 融合全局和序列特征的多变量时间序列预测方法. *计算机学报*, 2023, 46(1): 70-84)
- [2] Shao J, Mao Y, Zhang J. Task-oriented communication for multidevice cooperative edge inference. *IEEE Transactions on Wireless Communications*, 2022, 22(1): 73-87
- [3] Xie B, Wu Y, Shi Y, et al. Communication-efficient framework for distributed image semantic wireless transmission. *IEEE Internet of Things Journal*, 2023, 10(24): 22555-22568
- [4] Guo Y, Qin Z, Tao X, et al. Federated multi-view synthesizing for metaverse. *IEEE Journal on Selected Areas in Communications*, 2024, 42(4): 867-879
- [5] Tang S, Yang Q, Fan L, et al. Contrastive learning based semantic communications. *IEEE Transactions on Communications*, 2024, 72(10): 6328-6343
- [6] Dong C, Liang H, Xu X, et al. Semantic communication system based on semantic slice models propagation. *IEEE Journal on Selected Areas in Communications*, 2022, 41(1): 202-213
- [7] Zhang G, Hu Q, Cai Y, et al. SCAN: Semantic communication with adaptive channel feedback. *IEEE Transactions on Cognitive Communications and Networking*, 2024, 20(5): 1759-1773
- [8] Xie H, Qin Z. A lite distributed semantic communication system for Internet of Things. *IEEE Journal on Selected Areas in Communications*, 2020, 39(1): 142-153
- [9] Hu H, Zhu X, Zhou F, et al. One-to-many semantic communication systems: Design, implementation, performance evaluation. *IEEE Communications Letters*, 2022, 26(12): 2959-2963
- [10] Nan G, Liu X, Lyu X, et al. UDSem: A unified distributed learning framework for semantic communications over wireless networks. *IEEE Network*, 2023, 38(1): 296-303
- [11] Peng X, Qin Z, Tao X, et al. A robust semantic text communication system. *IEEE Transactions on Wireless Communications*, 2024, 23(9): 11372-11385
- [12] Xie H, Qin Z, Tao X, et al. Task-oriented multi-user semantic communications. *IEEE Journal on Selected Areas in Communications*, 2022, 40(9): 2584-2597
- [13] Zhang G, Hu Q, Qin Z, et al. A unified multi-task semantic communication system for multimodal data. *IEEE Transactions on Communications*, 2024, 72(7): 4101-4116
- [14] Wang C, Yu X, Xu L, et al. Multimodal semantic communication accelerated bidirectional caching for 6G MEC. *Future Generation Computer Systems*, 2023, 140: 225-237
- [15] Tong H, Yang Z, Wang S, et al. Federated learning based audio semantic communication over wireless networks//*Proceedings of the IEEE Global Communications Conference*. Madrid, Spain, 2021: 1-6
- [16] Weng Z, Qin Z, Tao X, et al. Deep learning enabled semantic communications with speech recognition and synthesis. *IEEE Transactions on Wireless Communications*, 2023, 22(9): 6227-6240
- [17] Wan Chen, Li Wen-Zhong, Ding Wang-Xiang, et al. A multivariate time series forecasting algorithm based on self-evolution and pre-training. *Chinese Journal of Computers*, 2022, 45(03): 513-525. (in Chinese)
(万晨, 李文中, 丁望祥等. 一种基于自演化预训练的多变量时间序列预测算法. *计算机学报*, 2022, 45(03): 513-525)
- [18] Zhou H, Zhang S, Peng J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting//*Proceedings of the AAAI Conference on Artificial Intelligence*. Virtual, 2021, 35(12): 11106-11115
- [19] Nie Y, Nguyen H, Sinthong P, et al. A time series is worth 64 words: Long-term forecasting with transformers//*International Conference on Learning Representations*. Kigali, Rwanda, 2023: 1-24

- [20] Zhang Y, Yan J. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting//The Eleventh International Conference on Learning Representations. Kigali, Rwanda, 2023: 1-21
- [21] Liu Y, Hu T, Zhang H, et al. ITransformer: Inverted transformers are effective for time series forecasting//The Twelfth International Conference on Learning Representations. Vienna, Austria, 2024: 1-25
- [22] Oreshkin N, Carpov D, Chapados N, et al. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting//International Conference on Learning Representations, Virtual Conference, 2020, 1-31
- [23] Wu H, Xu J, Wang J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting//Advances in Neural Information Processing Systems. Virtual Conference, 2021, 34: 22419-22430
- [24] Zeng A, Chen M, Zhang L, et al. Are transformers effective for time series forecasting? //Proceedings of the AAAI conference on artificial intelligence. Washington, USA, 2023, 37(9): 11121-11128
- [25] Wang H, Peng J, Huang F, et al. MICN: Multi-scale local and global context modeling for long-term series forecasting//The Eleventh International Conference on Learning Representations. Kigali, Rwanda, 2023: 1-22
- [26] Cao Shi-Xiang, Chen Chao-Meng, Tang Peng, et al. Differentially private federated learning with functional mechanism. Chinese Journal of Computers, 2023, 46(10): 2178-2195
(曹世翔, 陈超梦, 唐朋等. 基于函数机制的差分隐私联邦学习算法. 计算机学报, 2023, 46(10): 2178-2195)
- [27] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data//Artificial Intelligence and Statistics. Ft. Lauderdale, USA, 2017: 1273-1282
- [28] Li T, Sahu K, Zaheer M, et al. Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems, 2020, 2: 429-450
- [29] Karimireddy P, Kale S, Mohri M, et al. SCAFFOLD: Stochastic controlled averaging for federated learning//Proceedings of the International Conference on Machine Learning, Virtual, 2020: 5132-5143
- [30] Li X, Jiang M, Zhang X, et al. FedBN: Federated learning on non-IID features via local batch normalization//Proceedings of the International Conference on Learning Representations. Virtual, 2021: 1-27
- [31] Nguyen T, Torr P, Lim N. FedSR: A simple and effective domain generalization method for federated learning//Proceedings of the Advances in Neural Information Processing Systems. New Orleans, USA, 2022, 35: 38831-38843
- [32] Li D, Wang J. FedMD: Heterogenous federated learning via model distillation//Proceeding of the Neural Information Processing Systems, Vancouver, Canada, 2019:1-9
- [33] Huang W, Ye M, Du B. Learn from others and be yourself in heterogeneous federated learning//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022: 10143-10153
- [34] Makhija D, Han X, Ho N, et al. Architecture agnostic federated learning for neural networks//International Conference on Machine Learning. Baltimore, USA, 2022: 14860-14870
- [35] Collins L, Hassani H, Mokhtari A, et al. Exploiting shared representations for personalized federated learning//Proceedings of the International Conference on Machine Learning. Virtual Conference, 2021: 2089-2099
- [36] Alam S, Liu L, Yan M, et al. Fedrolex: Model-heterogeneous federated learning with rolling sub-model extraction//Proceedings of the Advances in Neural Information Processing Systems. New Orleans, USA, 2022, 35: 29677-29690
- [37] Tan Y, Long G, Liu L, et al. FedProto: Federated prototype learning across heterogeneous clients//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, Canada, 2022, 36(8): 8432-8440
- [38] Wu Wen-Tai, Wu Ying-Liang, Lin Wei-Wei, et al. Horizontal federated learning: research status, system applications and open challenges. Chinese Journal of Computers, 2025, 48(1): 35-67 (in Chinese)
(吴文泰, 吴应良, 林伟伟等. 横向联邦学习: 研究现状、系统应用与挑战. 计算机学报, 2025, 48(1): 35-67)
- [39] He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022: 16000-16009
- [40] Devlin J, Chang W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, USA, 2019, 1: 4171-4186
- [41] Su J, Ahmed M, Lu Y, et al. Roformer: Enhanced transformer with rotary position embedding. Neurocomputing, 2024, 568: 127063
- [42] Hu J, Shen L, Sun G. Squeeze-and-excitation networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 7132-7141



XING Huan-Lai, Ph. D. , associate professor, Ph. D. supervisor. His research interests include semantic communication, network function virtualization, software defined networks, artificial intelligence and evolutionary computation.

ZHAO Bo-Wen, Ph. D. candidate. His research interests include time series analysis, semantic communication, and deep reinforcement learning.

SONG Fu-Hong, Ph. D. His research interests include

mobile edge computing and multi-objective optimization.

XU Le-Xi, Ph. D. , professor-level senior engineer. His research interests include big data, network analysis and intelligent operation.

LEI Xian-Fu, Ph. D. , professor, Ph. D. supervisor. His research interests include B5G/6G wireless communications and next-generation Wi-Fi.

FENG Li, Ph. D. , Research professor, and Ph. D. supervisor. His research interests include cyberspace security and artificial intelligence.

Background

Traditional communication technologies focus on bit-level transmission accuracy, overlooking the inherent semantic features in source data, resulting in significant consumption of communication resources and weak real-time data analysis. Semantic communication emerges as one of the most promising techniques in 6G networks. Unlike traditional communication methods, semantic communication allows a transmitter to extract semantic information from source data and send task-oriented features to receivers. However, existing research mainly focuses on image and natural language data, which can usually be analyzed and processed using conventional algorithms such as CNN, RNN, and vanilla Transformer. In contrast, time series data normally exhibit periodicity, trends, and redundancy and contain noise interference. Existing research hardly considers these characteristics of time series data, leading to unnecessary consumption of communication resources and data processing efficiency reduction.

To achieve forecasting-oriented time series transmission, this paper constructs a heterogeneous federated semantic communication system for time series forecasting. By extracting and transmitting task-relevant semantic information of time series data from different devices, the efficiency of data transmission is improved, enabling joint inference among heterogeneous devices. To address the complex dependencies in time series data and enhance encoding efficiency, this paper designs a Multi-head Mixture Linear Attention

(MMA) time series encoding and decoding model to analyze the locality, periodicity, and redundancy characteristics of time series data. Additionally, a time series transmission scheme based on trend-seasonal decomposition is proposed to achieve robust time series transmission. To solve the problems of knowledge scarcity, low training efficiency, and poor inference performance among heterogeneous encoding models in distributed systems, this paper designs a Federated Embedding (FedEmb) algorithm to realize knowledge sharing among heterogeneous encoding models, ensuring the overall performance of the distributed semantic communication system. To improve training efficiency and system performance, this paper designs a two-stage pre-training paradigm for time series forecasting and feature reconstruction based on self-supervised learning in the semantic communication of time series data. It enhances the performance of time series semantic encoding and channel encoding models, thereby improving the accuracy of time series forecasting tasks at the receiver side. Simulation results prove that SC-FedEmb demonstrates excellent performance in time series forecasting tasks at the receiver side, surpassing six state-of-the-art baseline models with the lowest forecasting errors.

This work was supported by National Natural Science Foundation of China (No. U23A20273, No. 62271420), the Natural Science Foundation of Hebei Province (No. F2022105027), and the Fundamental Research Funds for the Central Universities, P. R. China.