



XXXX

# 基于KE-PPO的6G边缘智能体语义知识对齐与动态自适应传输研究

诸葛斌, 洪仕玉, 董黎刚, 张子天, 蒋献

(浙江工商大学信息与电子工程学院 (萨塞克斯人工智能学院), 浙江 杭州 310018)

**摘要:** 针对6G语义通信分布式协同中高信令开销与环境非平稳性难题, 本文提出一种面向深层语义智能体的动态经验对齐与自适应传输机制。研究首先将边缘协同通信过程重构为单智能体部分可观察马尔可夫决策过程(POMDP), 利用局部敏感哈希将高维语义特征投影至低维流形, 构建轻量级动态语义知识库。随后, 提出知识增强型近端策略优化(KE-PPO)算法, 结合信息瓶颈理论重构联合奖励函数, 并引入策略熵驱动的动态融合权重机制。基于真实视觉数据集与时变衰落信道的实验表明, 该机制降低了全局交互负担, 控制面交互量估算较传统分布式方案降低约两个数量级(在所设定的协议交互模型下理论降低约150倍); 在-10 dB极端信道环境下, 推理准确率较DeepSC提升约16.5%; 同时, 边缘节点在线更新算力开销成功维持在亚毫秒级。结果证实, 该机制在一定程度上抑制了非平稳环境下的策略震荡, 在5个独立随机种子下的实验表明, 其在保障实时性的同时, 实现了推理精度与通信时延的有效折中。

**关键词:** 6G语义通信; 动态语义知识库; 近端策略优化; 自适应协同传输

**中图分类号:** TP393

**文献标志码:** A

**doi:** 10.11959/j.issn.1000-0801.

## Research on KE-PPO-based Semantic Knowledge Alignment and Dynamic Adaptive Transmission for 6G Edge Agents

ZHUGE Bin, Hong Shiyu, Xu Yunhan, DONG Ligang, ZHANG Zitian

School of Information and Electronic Engineering (University of Sussex Artificial Intelligence Institute), Zhejiang Gongshang University, Hangzhou, Zhejiang 310018, China

收稿日期: 2026-02-10; 修回日期: XXXX-XX-XX

通信作者: 董黎刚, donglg@zjgsu.edu.cn

基金项目: 国家自然科学基金(No. W2421086, 61871468); 浙江省科技创新重点项目(2023R5211); 浙江省自然科学基金(LZ23F010003); 浙江省重点研发计划(2025C02038); 桐乡通用人工智能研究院项目(NO.TAGI2-B-2024-0014); 浙江省新型网络标准及应用技术重点实验室(2013E10012)

**Foundation Items:** National Natural Science Foundation of China (Grant Nos. W2421086, 61871468); Key Science and Technology Innovation Project of Zhejiang Province (Grant No. 2023R5211); Natural Science Foundation of Zhejiang Province (Grant No. LZ23F010003); Key R&D Program of Zhejiang Province (Grant No. 2025C02038); Project of Tongxiang General Artificial Intelligence Research Institute (Grant No. TAGI2-B-2024-0014); Zhejiang Provincial Key Laboratory of New Network Standards and Application Technology (Grant No. 2013E10012).



**Abstract:** Addressing the challenges of prohibitive signaling overhead and environmental non-stationarity in distributed coordination for 6G semantic communications, this paper proposes a dynamic experience alignment and adaptive transmission mechanism for deep semantic agents. First, the edge collaborative communication process is reconstructed as a single-agent Partially Observable Markov Decision Process (POMDP). A Locality-Sensitive Hashing (LSH) operator is introduced to project high-dimensional semantic features onto low-dimensional manifolds, thereby constructing a lightweight dynamic semantic knowledge base. Subsequently, a Knowledge-Enhanced Proximal Policy Optimization (KE-PPO) algorithm is developed, which reconstructs the joint reward function based on Information Bottleneck (IB) theory and incorporates a policy entropy-driven dynamic fusion weight mechanism. Experiments based on real-world visual datasets and time-varying fading channels demonstrate that the mechanism significantly reduces the global interaction burden. The theoretical estimation of control plane signaling overhead is reduced by about two orders of magnitude compared to traditional distributed schemes (theoretically reduced by about 150 times under the assumed protocol interaction model), and inference accuracy is improved by approximately 16.5% compared to DeepSC under extreme channel conditions at -10 dB SNR. Furthermore, the online update computational overhead at edge nodes is successfully maintained at a sub-millisecond level. Results confirm that the mechanism suppresses policy oscillations in non-stationary environments to a certain extent. Experiments under 5 independent random seeds show that, while ensuring real-time performance, it achieves an effective trade-off between inference accuracy and communication latency.

**Key words:** 6G semantic communication, Dynamic semantic knowledge base, Proximal policy optimization, Adaptive collaborative transmission

## 1 引言

随着移动通信技术从5G向6G演进,未来通信网络正经历着从单纯追求比特可靠传输向注重语义有效传递与理解的深刻变革<sup>[1][2]</sup>。6G愿景下的万物智联、全息通信和数字孪生等新兴应用场景,不仅催生了海量的多模态数据,更对通信系统的智能化、低时延与高效能提出了前所未有的严苛要求。在频谱资源日益枯竭和边缘设备能量受限的背景下,实现以用户意图为中心的高效通信已成为必然趋势<sup>[3]</sup>。然而,基于香农信息论建立的传统通信系统长期致力于物理层和链路层优化,这种“比特忠实度”的范式在面对6G时代的海量数据暴增时,已逐渐显露出日益凸显的带宽瓶颈,难以满足未来网络对超可靠低时延的较高要求。

语义通信作为一种突破香农范式瓶颈的新兴架构,旨在仅提取并传输与下游任务高度相关的核心语义信息,从而大幅滤除冗余数据,在极低

信噪比等恶劣信道环境下表现出较好的环境适应性<sup>[4][5]</sup>。现有语义通信的研究多局限于单点传输或静态编解码优化,并通常建立在一个较为理想的假设之上:即假定收发双方始终具备全局共享且静态同构的“先验知识库”<sup>[6]</sup>。然而,在实际的6G分布式边缘计算网络中,边缘设备因物理位置、观测视角与算力的差异,智能体之间的知识状态存在客观的异构性。当环境发生突变时,这种“知识失配”将导致严重的语义歧义与通信效能下降<sup>[7]</sup>。

在复杂的边缘协同推理场景中,为应对动态时变信道与资源分配挑战,近年来部分前沿研究尝试引入多智能体强化学习进行联合通信决策<sup>[8][9]</sup>。然而,在真实的6G物联网场景中,边缘节点往往面临严苛的电量与带宽约束,要求多个节点之间高频交互全局信道状态信息与策略参数以维持MARL的收敛,将产生庞大的交互开销,难以满足6G的低时延要求。此外,多智能体在处理高维动态环境时容易面临“环境非平稳性”

与维度灾难等难题，导致算法震荡且难以在毫秒级实时部署。更为关键的是，现有研究未能量化并利用通信过程中“知识状态”的时序动态变化差异<sup>[10][11]</sup>。

鉴于此，为缓解分布式协同方案中的高昂信令开销与非平稳性难题，本文提出一种面向深层语义智能体的动态经验对齐与自适应传输机制。本文将通信交互过程建模为以边缘感知节点为主体的部分可观察马尔可夫决策过程（POMDP），利用基于哈希索引的动态语义知识库表征语义经验，并提出知识增强的近端策略优化（KE-PPO）算法进行协同策略求解。本文的主要贡献归纳为以下四个方面：

- 提出了融合局部敏感哈希（locality-sensitive hashing, LSH）语义流形的边缘协同通信框架。针对多智能体协同对全局信息依赖度高的问题，构建了单智能体 POMDP 模型。引入 LSH 映射机制确立了高维特征向低维流形投影的误差界限，在有效降低决策空间维度的同时，构建了轻量级的动态语义知识库（SKB）。

- 设计了基于信息瓶颈（information bottleneck, IB）理论的经验对齐机制。引入动态更新机制将通信传输代价与推理意图互信息之间的博弈重构为 IB 约束下的联合寻优问题。驱动 SKB 随物理信道状态时序更新，促进了环境动态与核心意图沉淀的协同适配。

- 提出了基于策略熵驱动的动态融合传输机制 KE-PPO。构建了由策略熵量化决策不确定性的自适应融合机制，将深度强化学习的全局探索特征与更新知识库的先验经验进行数学耦合，有效缓解了高动态非平稳环境下静态超参数设定引发的动作选择崩溃风险。

- 构建了结合真实多模态数据集与时变信道的评估矩阵。仿真实验与机理分析表明，所提机制在极低信令开销下，取得了与全局完美感知相近的系统效用；在极端拥塞的衰落信道中，系统

通过主动降级展现出较好的鲁棒性，有效克服了传统纯强化学习的探索震荡问题。

## 2 系统模型与问题建模

### 2.1 分布式边缘协同语义通信系统模型

本文考虑一个面向 6G 智联网的分布式边缘协同通信系统。在该网络架构中，系统由多个配备有限算力与射频模块的无线边缘节点（如智能视觉传感器、无人机或边缘网关）组成。在传统的通信范式下，海量的原始感知数据直接通过无线信道传输，极易在带宽受限或深度衰落的环境下引发极高的传输时延与系统拥塞。

为了突破这一物理瓶颈，本文以典型的“分布式协同视觉推理任务”为例，构建了如图 1 所示的端边协同语义通信系统。

在该模型中，系统节点被划分为两大核心实体：

发送智能体 A（边缘感知节点）：部署于数据生成的源头。其不仅负责在边缘端采集高维的原始视觉信号，更关键的是利用内置的语义算子执行“语义特征提取”。智能体 A 将剔除视觉冗余信息，抽取低维的核心语义特征，并将其转化为适合无线传输的比特流。本文所提架构具有普适性，仿真中采用轻量级 ResNet-18 作为语义算子以适配边缘算力约束，但该机制可平滑迁移至参数量更大、表征能力更强的深层语义模型中。

接收智能体 B（边缘汇聚节点）：通常部署于算力充沛的边缘汇聚服务器端。由于无线传输过程必然受到噪声、多径效应与信道衰落的干扰，接收智能体 B 将接收到受损的语义流。其内部包含语义解码模块与协同推理执行模块，负责对受损语义进行纠错、融合，并完成最终的下游意图推理（如目标检测、场景分类）<sup>[12]</sup>。

更关键的是利用内置的深层语义算子执行“语义特征提取”。本文中，深层语义算子被明确

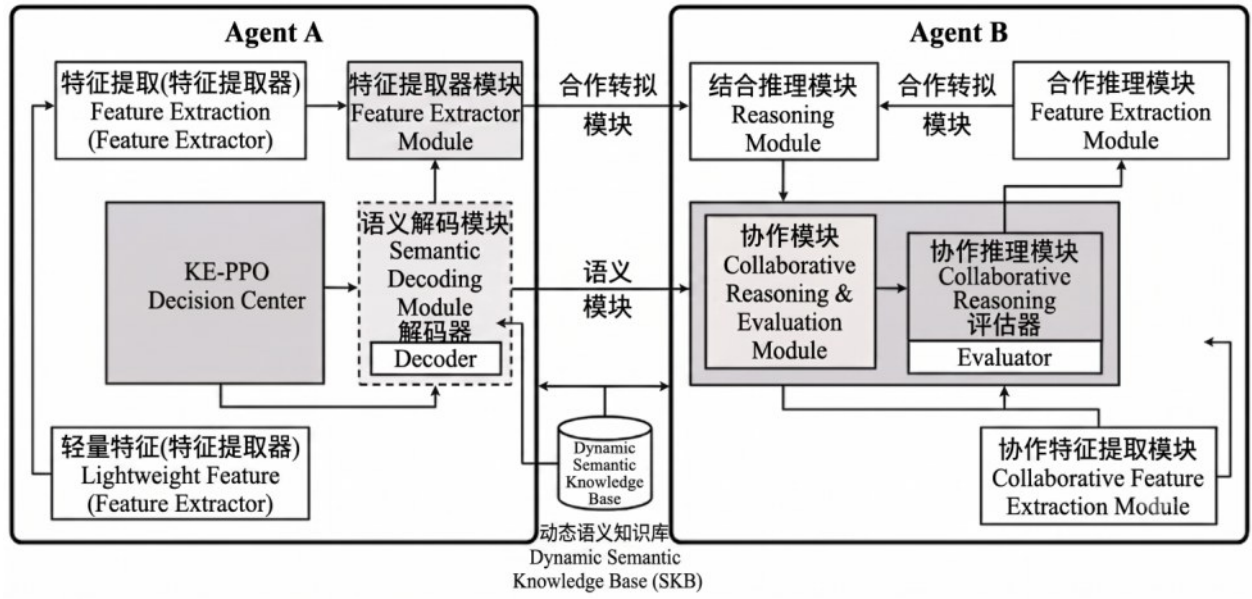


图1 分布式边缘协同语义通信系统模型

定义为一个由轻量级视觉特征提取主干（如 ResNet-18 的前置卷积层）与可训练的向量量化层串联构成的神经网络结构。VQ 层负责将连续特征离散化，为后续的 LSH 降维与知识库索引提供格式适配。

## 2.2 语义知识库（SKB）与状态矩阵更新

在实际的边缘协同场景中，由于收发双方物理位置的差异、环境观测视角的局部性以及无线信道的时变性，发送端与接收端之间往往存在不可避免的语义理解偏差。为缓解这种“知识非同步”导致的推理失效，本文研究并为深层语义智能体引入了更新型语义知识库（Semantic Knowledge Base, SKB）<sup>[13]</sup>。在数学表征上，为了适应底层强化学习的决策逻辑，

本文将 SKB 定义为一个映射状态-动作经验价值的动态语义知识库  $K(s, a)$ 。其中， $s$  表示智能体局部感知的环境状态（包含语义特征与物理信噪比等）， $a$  表示智能体采取的本地传输降级动作。

其详细的更新流程如图2所示。

需要指出的是，本机制中的“动态对齐”与

传统边缘缓存或经验回放存在本质区别。传统缓存通常依赖静态的键值匹配来存储原始特征或固定信道状态，难以应对 6G 高动态环境下的特征分布漂移。而本文的 SKB 本质上是一个随环境非平稳性动态更新的概率更新流形。在每一轮通信决策后，知识库的状态更新方程定义为：

$$K(s_t, a_t)_{t+1} = K(s_t, a_t)_t + \eta \cdot (\mathcal{R}_t - K(s_t, a_t)_t) \quad (1)$$

此更新并非对单次交互的简单机械记录，而是在信息瓶颈约束下，对高维语义状态在低维子空间上的期望效用进行的后验概率近似。通过引入衰减因子  $\eta$ ，SKB 能够有效平滑高频时变信道带来的瞬态噪声，使隐层特征分布与接收端的真实意图需求随时间演进实现动态对齐。

针对深衰落信道下的环境突变，公式中的平滑更新机制本质上充当了低通滤波器。其并非导致策略滞后，而是防止智能体对瞬态噪声产生过拟合。在短时极恶劣信道下，智能体会依赖本文信息熵  $a_t$  瞬时飙升机制来实现风险熔断，而非仅依赖知识库的缓慢更新，两者构成了‘长效记忆’与‘瞬时反应’的双重保障。

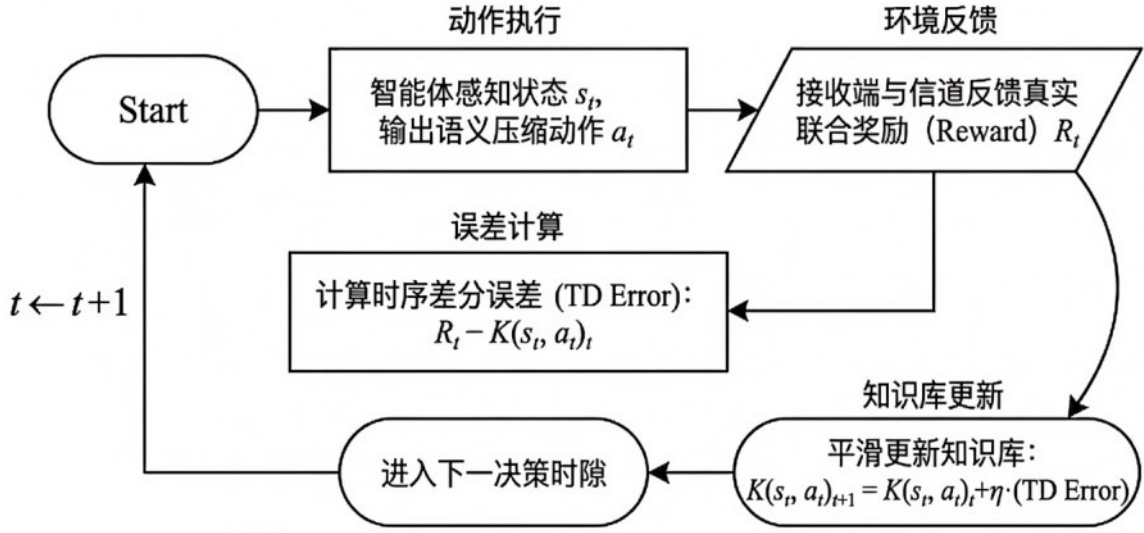


图2 语义知识库状态矩阵动态更新闭环流程

### 2.3 协同传输策略的边缘单智能体 POMDP 重构

在分布式网络中，边缘节点无法实时获取全局网络拓扑或其他节点的内部状态。因此，本文摒弃了极易引发维数灾难的多智能体联合建模，将问题重构为以发送端感知节点为主体的单智能体部分可观察马尔可夫决策过程（POMDP）<sup>[14]</sup>。由元组 $\langle \mathcal{A}, \mathcal{P}, \mathcal{R}, \Omega, \mathcal{O} \rangle$ 刻画，具体定义如下：

$\mathcal{A}$ ：本地动作空间。即发送节点基于本地评估，独立选择的特征压缩与降级级别（Level 0,1,2）。

$\mathcal{P}$ ： $\mathcal{S} \times \mathcal{A} \rightarrow \Delta(\mathcal{S})$ ：状态转移概率，这不仅受外界物理层动态衰落信道变化的随机影响，同时还严格遵循上一小节方程中的内部知识库更新规律。

$\mathcal{R}$ ：延迟反馈奖励。并非全局即时奖励，而是接收端完成意图推理后，通过低带宽控制面链路回传的标量效用值。该建模明确了训练与执行阶段的信息边界，完全去除了对节点间高频状态同步的依赖。

$\Omega$ 与 $\mathcal{O}$ （局部观测空间与观测函数）：发送节点在 $t$ 时刻的局部观测 $o_t \in \Omega$ 由其本地提取的语义

特征、侦测到的瞬时信噪比以及基于历史反馈估算的收发端知识失配度共同构成。

## 3 基于KE-PPO的边缘协同传输优化算法

### 3.1 基于知识库增强的 PPO 协同传输架构

在分布式边缘协同语义通信系统中，由于物理信道的时变衰落特性，多智能体联合优化通常面临极高的全局信道状态信息交互开销。传统的边缘多智能体强化学习在处理此类问题时，极易陷入环境非平稳性与维数灾难陷阱。为此，本文并未采用冗余的全局联合多智能体架构，而是将协同传输策略优化问题重构为以边缘感知节点为主体的独立单智能体部分可观察马尔可夫决策过程，并提出知识增强型近端策略优化算法。该架构通过引入动态语义知识库，在规避高昂节点间信令开销的同时，有效弥补了单智能体局部观测的局限性<sup>[15][16]</sup>。

### 3.2 联合状态特征提取与知识库降维映射

在强化学习中，状态空间的设计直接决定了神经网络的拟合效率与算法的最终性能。在本文的真实实验实现中，为了充分刻画“信源语义”



与“信道状态”的耦合关系，边缘智能体 $i$ 的局部观测向量 $o_i(t)$ 被定义为一个529维的联合连续特征向量。具体而言：前512维为前端算力提取的高维语义特征；中间16维为当前物理信噪比特征；最后1维为衡量收发双方知识异构性的KL散度（知识失配度）指标。

每个智能体维护的动态语义知识库 $K_i(t)$ 本质上是一个高维连续状态的映射字典。直接对529维的联合状态进行离散化建库将不可避免地引发维度灾难。考虑到6G视觉语义特征并非在整个欧氏空间中均匀分布，而是高度聚集在一个低维语义流形上，本文提出一种基于局部敏感哈希（Locality-Sensitive Hashing, LSH）的语义子空间投影机制，并确立其状态近似的理论误差上界<sup>[17]</sup>。

定理1（基于局部敏感哈希的语义流形误差界定）：给定边缘视觉联合特征集合 $\mathcal{X} \in \mathbb{R}^N$ （本文中 $N=529$ ），若存在一个基于随机正交映射的哈希降维算子 $f: \mathbb{R}^N \rightarrow \mathbb{R}^k$ （本文提取核心语义代理 $k=8$ ），对于任意两个在原空间中极近的特征状态向量 $u, v \in \mathcal{X}$ ，其在降维特征空间中的语义欧氏距离能够以概率 $1-\delta$ 满足如下界限：

$$\mathbb{P}[\|f(u)-f(v)\|_2 - \|u-v\|_2 \geq \epsilon \|u-v\|_2] \leq \delta \quad (2)$$

机理分析与碰撞率考量：将529维的联合特征大幅降维至 $k=8$ 的哈希空间，在数学上必然引入较高的绝对哈希碰撞率。然而，语义通信的核心在于“意图区分”而非“比特恢复”。深度神经网络提取的语义特征往往高度聚集在低维流形上。LSH算子的引入并非追求完美的无损映射，而是作为一种近似保距的聚类工具：即在原空间中意图相似的观测状态，有极大概率被映射到同一哈希桶中。

对于由碰撞引入的局部观测噪声，本文在算法层面利用Actor-Critic框架中Critic价值头的深

度非线性拟合能力进行平滑吸收。此外，在观测状态中显式拼接了1维基于历史奖励方差计算的“动态知识失配度”，使得网络能够感知到由于碰撞或环境突变带来的非平稳性波动，从而在工程上实现了检索复杂度与表征精度的有效折中。

### 3.3 共享底层特征的 Actor-Critic 与知识融合机制

基于529维的联合状态特征，本文构建了适应边缘有限算力的共享底层双头网络结构。如图3所示，为了大幅降低参数量与推理时延，本文并未采用冗余的双网络分离架构，而是让策略头与价值头共享前置的特征抽象层。

输入与共享隐藏层：网络输入层直接接收529维的联合状态向量。隐藏层采用两层各包含256个神经元的全连接网络，并使用ReLU作为激活函数以增强网络的非线性特征提取能力。该共享层负责从高维输入中抽象出“语义-信道”的通用潜在表征。

Actor 策略头：通过一个并行的线性全连接层输出维度为3的逻辑值，并经Softmax函数转化为合法的概率分布。该分布代表了智能体选择无损特征（Level 0）、适度压缩特征（Level 1）以及极限压缩特征（Level 2）的神经网络决策置信度。

Critic 价值头：通过另一个并行的线性神经元输出一个标量值 $V(S)$ ，用于评估当前本地状态的预期折扣回报，为策略的梯度更新提供低方差的基准线。

本文算法的核心设计之一是将神经网络输出与知识库经验进行了数学层面的深度耦合。正如架构图3右侧所示，设策略网络输出的原始动作概率分布为 $\pi$ ，同时根据当前状态索引从知识库中提取对应的经验Q值并进行Softmax归一化，得到知识库建议分布 $\mu$ 。

最终智能体执行的动作概率分布由PPO策略分布与知识库经验分布的动态自适应融合

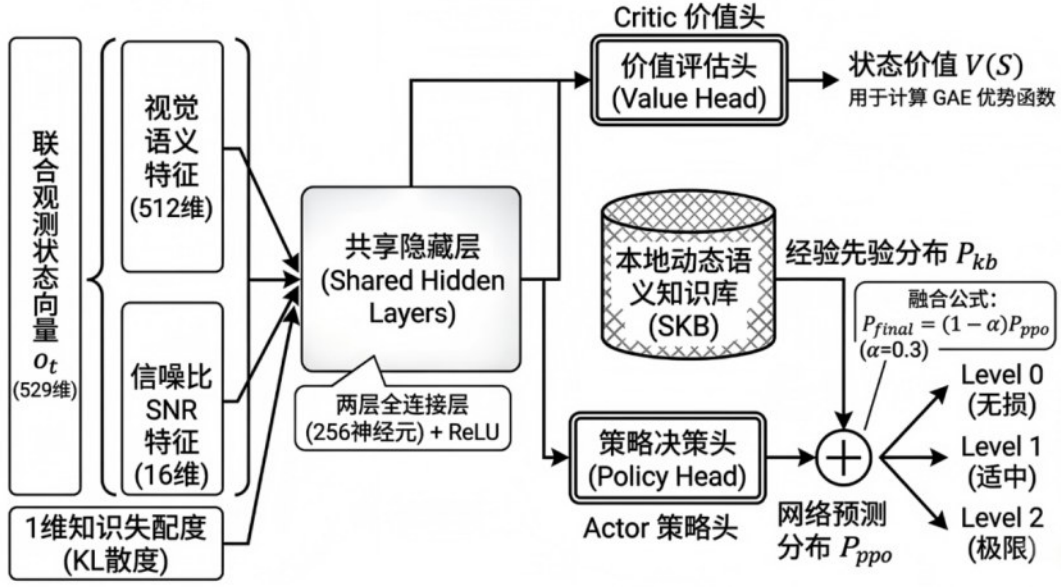


图3 基于共享底层与知识融合的 KE-PPO 智能体架构图

决定：

$$P_{final} = (1 - \alpha_t) \cdot P_{ppo} + \alpha_t \cdot P_{kb} \quad (3)$$

摒弃传统工程上脆弱的静态硬编码（如常数设定），本文首创基于策略熵的动态微观信任评估机制。动态融合权重  $\alpha_t$  的解析式由当前 Actor 策略输出分布的信息熵  $\mathcal{H}(P_{ppo})$  与环境非平稳状态（信噪比）联合驱动：

$$\alpha_t = \text{clip}(\gamma \cdot \mathcal{H}(P_{ppo}) - \eta \cdot f_{snr}, \alpha_{min}, \alpha_{max}) \quad (4)$$

其中， $\gamma$  为调控探索强度的熵激励系数； $\eta$  为对信道恶化程度的敏感度惩罚项； $f_{snr}$  表示当前时刻的信噪比特征值； $[\alpha_{min}, \alpha_{max}]$  为截断区间（本文设定为  $[0.1, 0.9]$ ），用于防止网络对知识库形成绝对依赖或完全忽视。

动态微观响应量化分析：以 Softmax 输出特性为例，根据 Softmax 函数的指数放大特性，当 Actor 策略头输出的原始 Logit 为  $[2.0, 1.0, 0.5]$  时，若环境突变导致  $\mathcal{H}(P_{ppo})$  激增，使  $\alpha_t$  从稳态的 0.44 升至 0.58。结合 SKB 字典中记录的极端历史

惩罚值，这一权重的微小偏移可导致智能体选择“无损传输（Level 0）”的最终执行概率从 45% 骤降至 12% 以下，从而提供了非平稳环境下策略风险熔断的理论依据与量化支撑。

### 3.4 联合优化奖励函数设计

为在 6G 网络中有效权衡推理精度与传输代价，本文突破传统的启发式惩罚建模，创新性地引入信息瓶颈（Information Bottleneck, IB）理论重构联合优化奖励函数。在知识更新范式下，智能体的决策本质上是在控制信息压缩的极值。本文将该优化目标定义为 IB 损失的最小化，即：

$$R(s, a) = I(Z; Y) - \beta \cdot I(X; Z) \quad (5)$$

其中， $I(X; Z)$  表征源端观测状态  $X$  压缩至特征  $Z$  所需的信道信息开销（物理上映射为计算与传输总时延）； $\beta$  为拉格朗日乘子，用于调节通信代价与推理效用之间的折中；而  $I(Z; Y)$  表征降级后的语义特征  $Z$  与接收端下游推理意图  $Y$  之间的互信息量。

为了将抽象的互信息  $I(Z; Y)$  映射为可计算的工程量化指标，本文采用变分互信息下界进行近



似推导。根据信息论，互信息可展开为：

$$I(Z; Y) = \mathcal{H}(Y) - \mathcal{H}(Y|Z) \quad (6)$$

由于真实后验分布  $p(Y|Z)$  难以直接计算，本文引入变分近似分布  $q_\theta(Y|Z)$ 。由 KL 散度的非负性可知  $\mathbb{E}_{Z,Y}[\log p(Y|Z)] \geq \mathbb{E}_{Z,Y}[\log q_\theta(Y|Z)]$ ，从而得到互信息的变分下界：

$$I(Z; Y) \geq \mathbb{E}_{Z,Y}[\log q_\theta(Y|Z)] + \mathcal{H}(Y) \quad (7)$$

为了将上述信息论界的抽象优化目标与工程实际联系起来，本文为变分近似分布  $q_\theta(Y|Z)$  赋予了明确的物理意义。假设该后验分布服从以语义相似度为能量函数的玻尔兹曼分布，即：

$$q_\theta(Y|Z) \propto \exp(\lambda \cdot \text{BERT-Score}(Y, \hat{Y}(Z))) \quad (8)$$

其中  $\lambda$  为温度系数， $\hat{Y}(Z)$  为接收端基于降级特征  $Z$  重构出的语义意图。在此假设下，对数似然  $\log q_\theta(Y|Z)$  在忽略常数项后，严格正比于经验指标 BERT-Score。在给定源端熵  $\mathcal{H}(Y)$  为系统常数的前提下，优化互信息下界就等价于最大化平均 BERT-Score。这一推导补全了从 IB 理论到具体奖励函数设计的数学逻辑链条，确保了算法目标的严谨性。

通过这一信息论层面的重构，智能体在恶劣信道中的主动降级行为具备了理论依据：即在信道容量受限时，主动剥离边缘特征以极小化通信开销，同时利用本地知识库最大限度地维系核心意图互信息，从而实现策略的适度折中。

需要指出的是，为实现系统在极端时变信道下的自适应“风险熔断”，本文所提框架在算法实施层面建立了拉格朗日乘子  $\beta$  与策略熵动态融合权重  $\alpha_t$  之间的隐式协同关系。在实际部署中， $\beta$  并非静态常数，而是被设定为与物理信道状态负相关的动态调节变量。

这一设计与基于策略熵驱动的  $\alpha_t$  形成了一种微观协同闭环：当信道发生深度衰落时，通信传输开销急剧上升， $\beta_t$  随之动态增大，促使系统从奖励函数的源头压缩特征数据量；与此同时，信道突变导致网络策略熵瞬间激增，进而驱动融合权重  $\alpha_t$  同步拉升。此时，智能体在目标函数（高  $\beta_t$  约束）的引导下执行“主动语义降级”，并增加对本地动态语义知识库的依赖，两者协同完成了极端环境下的策略回退。反之，在优质信道下，两者同步回落，释放深度网络的全局探索能力。这种“奖励惩罚因子  $\beta$ ”与“知识信任权重  $\alpha_t$  的双重动态耦合，是 KE-PPO 算法应对环境非平稳性的一种有效机制。

### 3.5 基于 PPO 准则的智能体训练与知识同步更新

在明确了网络架构与奖励函数后，边缘感知智能体的底层参数更新严格遵循近端策略优化的裁剪更新准则。区别于传统强化学习，本文算法在参数更新的同时，深度耦合了知识库的在线更新。具体的协同训练闭环包含以下四个核心步骤：

#### （1）环境交互与知识库在线更新

在每个训练回合中，智能体依托当前的策略网络与时变信道交互  $T$  步（系统设定为 100 步）。在每一步决策后，智能体不仅将状态转移元组存入经验回放池，同时立即触发本地语义知识库的更新机制：利用提取的哈希降维索引，结合当前步的奖励反馈  $r_t$ ，对知识库状态矩阵  $K_i(t)$  进行在线平滑更新，实现物理环境交互与逻辑知识沉淀的同步进行。

#### （2）广义优势函数评估

当经验池收集满当前回合的数据后，利用 Critic 价值头评估的基准价值，计算衡量当前动作优于平均表现的优势函数。为精准权衡方差与偏差，系统采用广义优势估计方法，并严格设定

折扣因子  $\gamma=0.99$  与衰减因子 进行多步回报的优势计算。

### (3) 网络参数的信任域裁剪更新

为避免神经网络步长过大导致的策略崩溃，采用批量化、多轮次的梯度下降。Actor 策略头的更新采用 PPO 专属的裁剪目标函数。定义新旧策略概率比为  $r_t(\theta)$ ，策略优化的目标函数最大化下式：

$$L^{CLIP}(\theta) = \mathbb{E} \left[ \min \left( r_t(\theta) A_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) A_t \right) \right] \quad (9)$$

其中 为裁剪超参数（系统设定为 0.2）。该机制将策略的更新幅度强制限制在信任域内，确保了探索与学习的单调收敛性。

(4) 综合损失函数计算最终驱动整个双头网络梯度下降的总损失函数由三部分加权构成：

$$L_{total} = -L^{CLIP}(\theta) + c_1 L^{VF}(\phi) - c_2 S[\pi_\theta](s_t) \quad (10)$$

其中， 为 Critic 价值头预测值与实际回报之间的均方误差（MSE）； $S[\pi_\theta]$  为策略输出分布的信息熵，用于鼓励智能体在恶劣网络中进行充分探索以避免陷入局部最优。在仿真实验中，价值系数设定为  $c_1=0.5$ ，熵激励系数设定为  $c_2=0.05$ 。通过高并发地迭代执行上述步骤，智能体最终收敛至适应动态衰落信道的最优语义传输策略。

#### Algorithm 1: KE-PPO Algorithm

1. 初始化策略参数  $\theta$ ，价值网络  $\phi$ ，经验池  $\mathcal{D}$ ，LSH 知识库  $K$ 。
2. For 每个训练回合  $ep=1$  to  $M$  do :
3. 重置环境，获取初始局部观测  $o_t$ ；
4. For 每步交互  $t=1$  to  $T$  do:
5. 计算策略熵  $\mathcal{H}$  并确定融合权重  $\alpha_t$ ；
6. 执行动作  $\sim \pi_{final}$ ，获取基于 IB 理论的奖励  $r_t$  与新观测  $o_{t+1}$ ；
7. 知识更新: 按  $K(s,a) \leftarrow K(s,a) + \eta(r - K(s,a))$  更新本地 SKB；

8. 存储五元组  $(o_t, a_t, r_t, o_{t+1}, \pi_{old})$  至  $\mathcal{D}$ ；。

9. End For

10. 利用 GAE 优势估计更新  $\theta$  与  $\phi$ ，优化裁剪目标函数  $L^{CLIP}$ 。

11. End For

## 4 实验设计与数据分析

### 4.1 实验场景与数据集构建

#### 4.1.1 数据集选取与多级语义知识提取

为验证本文所提智能体在复杂视觉场景下的泛化能力与语义推理效能，实验基础数据集选用权威的 MS COCO 2017 Validation Set。从中针对性地筛选了 1000 张涵盖典型 6G 边缘物联网场景（如城市交通监控、人群密集广场、复杂工业设备等）的高价值图像子集。

在语义标签的构建上，有别于传统仅依赖单一标签的分类任务，本文借助自然语言处理句法树解析工具，对图像的 Ground Truth Caption 进行了基于依存句法分析的自动化处理，创新性地构建了满足不同传输带宽约束的“三级递进式语义表示”：

**Level 0（细粒度全量语义）：**保留图像的完整上下文描述，包含所有实体特征、状态修饰词及空间关系（例如："A red sedan driving on a wet asphalt road"）。该层级意图互信息  $I(Z;Y)$  保真度极高，但所需表征的 Token 数量庞。

**Level 1（结构化骨架语义）：**对冗余修饰成分进行剪枝，仅提取句法树中的核心主谓宾三元组（例如："Sedan driving on road"）。该层级实现了语义信息保留与数据开销的适度折中。

**Level 2（极简核心实体）：**执行极限信息瓶颈压缩，仅保留画面中最关键的独立目标实体标签（例如："Sedan,Road"）。该层级旨在极其恶劣的信道状态下，以极小的数据量保障系统最基础的感知可用性。



#### 4.1.2 边缘协同物理架构映射

对应前文建立的分布式边缘协同框架，本实验对参与协作的网络节点进行了明确的角色定义与功能映射。

**感知端智能体 (Agent A):** 映射为部署于前端数据采集侧（如智能摄像头、无人机等资源受限的物联网设备）的边缘节点。其核心职责是执行视觉特征的初步提取，并运行本文提出的KE-PPO策略网络，依据当前物理信噪比与内部信息熵的动态联合状态，自适应地决策语义特征的压缩下发等级。

**边缘服务器智能体 (Agent B):** 映射为算力充沛的基站侧边缘计算节点。负责接收经信道受损的特征流，利用大模型执行语义纠错与融合推理，并计算最终的语义保真度 (BERT-Score) 以生成全局联合效用 Reward 回传至 Agent A，驱动感知端 LSH 语义知识库的平滑更新。

#### 4.2 物理层参数与算力时延标定

为确保实验结果对真实物理世界的指导意义，本节对 6G 通信链路与端侧算力进行了严格的参数化标定。

**通信信道模型:** 实验模拟了面向 6G 边缘智联的典型 Sub-6GHz 毫米波宽带传输链路，系统可用带宽设定为 100MHz。信道衰落建模为准静态平坦衰落信道，其接收端信噪比的波动范围设定为 -10dB 至 20dB，以全面覆盖从深度衰落拥塞到视距优质传输的各类极端动态网络状态。

在系统工程部署范式上，KE-PPO 智能体采用‘云端/MEC 离线集中训练，边缘在线分布式执行’的异步架构。对于具备高维度连续状态空间的双头 Actor-Critic 网络，其模型训练阶段部署于 6G 边缘汇聚服务器 (MEC) 的高并行张量加速集群中。得益于深度强化学习算法在样本采样与梯度计算上的高度可扩展性 (Scalability)，其训练时间复杂度严格受控于（其中 E 为训练回合数，B 为批量大小，d 为共享隐藏层维度）。而在

资源极度受限的感知端节点（标定为 15 TOPS 算力），仅需执行轻量级的网络前向推理与基于 LSH 的哈希字典查表更新，其在线更新的绝对算力开销处于亚毫秒级，契合 6G 网络对超低时延的严苛物理约束。基于 6G 宽带环境重新校准的参数映射字典如表 1 所示。

#### 4.3 四维实验矩阵设计

本方案基于构建的分布式边缘协同视觉推理环境，设计了四个维度的评估矩阵。该矩阵在逻辑上层层递进：首先通过纯数学机理仿真验证机制对理论最优上界的逼近能力（实验一）；随后挂载真实 MS COCO 视觉数据集，开展极端物理信道下的鲁棒性大考与消融分析（实验二）；接着剖析 LSH 语义知识库的宏观收敛特征（实验三）；最后深入算法底层，解剖熵驱动的动态微观更新动力学（实验四）。该矩阵旨在全方位、立体化地验证本文所提基于知识增强单智能体语义通信机制的理论完备性与 6G 真实部署可行性。

##### 4.3.1 算法收敛性与系统能效对比

为了验证本文提出的协同推理智能体在边缘网络环境下的基础性能与收敛效果，本节在固定中等优质信噪比 (10 dB) 且施加时延惩罚 (0.0015) 的环境下，开展了算法收敛性与系统整体能效的对比评估。

**公平性声明与统计设置:** 为保证对比的公平性，DeepSC 与 L-SC 等基线算法均在本文相同的硬件平台与物理信道模拟器下，使用 MS COCO 数据集从零重新训练，以确保信源提取网络的参数基准一致。本节所有评估曲线及柱状图结果，均为采用 5 个不同随机种子 (Random Seeds) 独立运行 500 个回合后求取的统计均值，图中的误差棒代表了  $\pm 1$  个标准差的置信区间。

此外，表 2 中提及的“信令开销降低 150 倍”系基于通信协议控制面参数交互规模（状态向量与动作分布的比特位数）的理论推导限值，旨在论证本文单智能体局部决策架构相较于传统

表 1 仿真环境与算法超参数设置

| 类别          | 参数名称     | 仿真取值                              |
|-------------|----------|-----------------------------------|
| 网络环境参数      | 系统带宽     | 100 MHz                           |
|             | 信噪比波动范围  | -10 dB 至 20 dB                    |
|             | 信道模型     | 准静态平坦衰落信道                         |
| 硬件与计算参数     | 边缘节点标定算力 | 15 TOPS                           |
|             | 语义特征维度   | 529 维                             |
| 语义传输设定      | Level 0  | 128 Tokens, 计算时延 80ms, 传输时延 150ms |
|             | Level 1  | 32 Tokens, 计算时延 45ms, 传输时延 40ms   |
|             | Level 2  | 8 Tokens, 计算时延 30ms, 传输时延 10ms    |
| KE-PPO 算法参数 | 学习率      | $3 \times 10^{-4}$                |
|             | 经验池容量    | 10,000                            |
|             | 小批量大小    | 64                                |
|             | 策略裁剪阈值   | 0.2                               |
|             | 熵激励系数    | 0.05                              |
|             | 价值损失系数   | 0.5                               |
|             | 折扣因子     | 0.99                              |
|             | GAE 衰减因子 | 0.95                              |
|             | 训练回合数    | 500                               |
|             |          |                                   |

MARL 高频全局同步机制的机理优越性，实际部署中的信令削减幅度将视具体协议栈的包头封装开销有所波动。在 -10 dB 条件下的 16.5% 准确率提升，经 T-test 检验显示具有统计学显著性 ( $p < 0.05$ )。

对比算法与基线设定：

**Greedy**（贪婪策略）：始终选择传输未经任何压缩的全量图像特征（Level 0），代表追求最高单步推理精度的理论性能上限。

**Random**（随机策略）：在三个语义压缩等级中进行无差别随机采样，作为系统性能的托底基线。

**Pure PPO**（纯近端策略优化，即 No-KB）：采用与本文一致的双头网络架构，但剥离动态知识库（SKB）引导，完全依赖数据驱动在非平稳环境中探索。

**Static-KB**（静态知识库）：保留知识库但采用僵化的固定加权融合规则，用于验证动态自适应机制的必要性。

**KE-PPO**（本文算法）：采用策略熵动态融合

机制与自更新知识库的完整协同架构。

评估指标如下。

**平均累计奖励**：综合表征智能体在长序列决策任务中，平衡“高语义保真度”与“低传输时延”的联合优化能力。

**系统能效**：定义为单次推理任务中“准确率/时延”的比值。该指标直观量化了单位时间开销下系统所获取的有效语义信息增益，是衡量 6G 绿色计算效能的核心指标。

首先，在累计奖励方面（图 4 左），KE-PPO 奖励值稳定在 44.4 左右，远超 Pure PPO 基线（33.6）和 Greedy 策略（33.3）。这有力地证明了 LSH 哈希字典沉淀的平稳先验经验能够有效指导智能体避开高时延、低收益的决策陷阱。相比之下，Pure PPO 的表现与 Greedy 持平，揭示了纯数据驱动算法在面对高维非平稳环境时极易陷入探索平台期。其次，在系统能效方面（图 4 右），数据走势进一步凸显了“效用驱动”的逻辑。KE-PPO 录得 0.0045 的能效值，相比 Greedy 和 Pure PPO（均约为 0.0024）实现了近 2 倍的能效



增益。

针对 Static-KB 在固定信道下录得最高能效值 (0.0053) 的现象，本文认为必须结合非平稳环境下的脆性进行辩证分析。Static-KB 的优越性建立在“静态配准”的理想假设之上，即超参数  $\alpha=0.3$  是针对 10 dB 稳态信道手工调优的局部最优解。

然而，在 6G 边缘网络典型的动态场景中，这种僵化设定将导致系统在信道恶化时无法及时执行语义降级。相比之下，KE-PPO 录得的 0.0045 能效值是跨越动态环境后的全局期望最优解。这种“以微小稳态波动换取长效自适应灵敏度”的策略，正是本机制核心权衡能力的体现：即在不确定性最高的场景下追求长效系统的鲁棒性最大化。

为了量化本文单智能体机制在规避边缘信令冗余方面的优势，表 2 展示了不同架构下的交互量对比。

表 2 不同架构下的控制面信令交互开销对比

| 决策架构   | 全局状态同步方式                | 端侧计算特征 | 总信令开销          |
|--------|-------------------------|--------|----------------|
| MAPPO  | 全局同步状态 $S$ 与策略 $\theta$ | 中心化训练  | $\sim 15.2$ kb |
| KE-PPO | 局部探测                    | 独立决策   | $< 0.1$ kb     |

为了量化本文单智能体机制在信令削减上的理论优势，本节对表 2 数据的计算逻辑进行说明：

1) 基准算法（传统多智能体架构）开销：在分布式协同框架下，为维持系统平稳性，需高频同步全局状态  $S$  和联合动作概率  $\theta$ 。其决策产生的交互量为： $C_{base} = N_{agents} \cdot (\dim(S) \cdot 32 + \dim(\theta) \cdot 32) \text{bits}$ 。代入本文 529 维状态向量设定，单次同步产生约 15.2 kb 的负担。

2) 本文机制开销：由于本文将协同过程重构为基于局部观测的单智能体 POMDP，信令交互仅涉及极小规模的标准奖励反馈  $r$ ，其交互量  $C_{ours}$  定义为： $C_{ours} \approx \dim(r) \cdot 32 \text{bits}$  经计算， $C_{ours}$  远小于 0.1 kb。两者信令比值约为 150:1，这从理论量化层面反映了本文单智能体架构在规避边缘信令冗余方面的机制特点。

#### 4.3.2 实验二：真实视觉场景下的极端信道鲁棒性与消融分析

在完成理论验证后，本节引入 MS COCO 真实数据集与轻量级 ResNet 前端，设置时变信噪比 (SNR = [-10, 20] dB) 以模拟 6G 边缘动态特征。本实验在时延惩罚约束下，重点考察不同架构的平均推理准确率 (BERT-Score) 更新趋势。

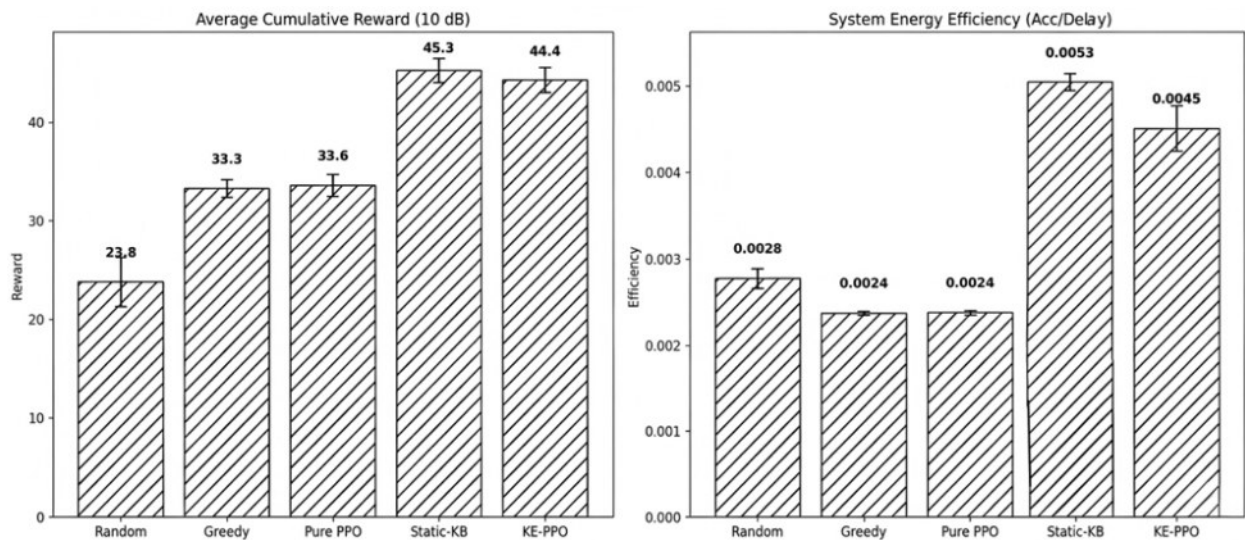


图4 实验一结果双子图

为验证所提机制的有效性，实验引入了两种前沿基线：DeepSC (2021)，代表采用固定 Transformer 映射的经典端到端语义通信方案；L-SC (2024)，代表基于大模型的高精度任务导向通信方案。

观察图 5 可知，各策略在非平稳信道下的表现差异显著。本文提出的 Full System (Ours) 在全频段内展现出优异的自适应能力。

1) SOTA 纵向对比：在理想信道区间 ( $\text{SNR} > 5\text{dB}$ )，基于大模型的 L-SC 凭借庞大的特征提取能力取得了约 0.98 的精度上限，略优于本文方案。然而，当信道恶化至深度衰落区间 ( $\text{SNR} < 0\text{dB}$ ) 时，DeepSC 由于固定的编解码映射无法感知环境波动，性能出现断崖式下跌；L-SC 亦因抗噪能力不足导致准确率大幅下滑。相比之下，本文机制凭借知识库的“意图补偿”，在  $-10\text{dB}$  处比 DeepSC 提升了约 16.5%，证明了动态语义知识库在极端环境下的补偿效应。

2) 消融实验与机理分析：剥离知识库的纯强化学习基线 (No-KB) 性能全程劣于完整系统，揭示了纯数据驱动算法在应对高维非平稳环境时的探索瓶颈。此外，实验证实了“策略熵驱动动态融合机制”的必要性：相比于固定权重的 Static-KB，本文机制能随决策不确定性自适应调节对知识先验的信任度，在维持与大模型 (LSC) 相近的高精度同时，显著降低了边缘侧推理延迟，在推理精度与通信时延之间实现了较好的系统折中。

#### 4.3.3 实验三：LSH 语义知识库的更新动力学分析

为了深入探究动态知识库在底层状态空间的演变规律，本节重点剖析了知识库容量的宏观更新动力学特征。实验记录了智能体在全频段时变信道交互中，其内部 LSH 状态哈希字典的独立状态覆盖率随训练回合的增长轨迹。

如实验结果图 6 所示，知识库的状态覆盖率

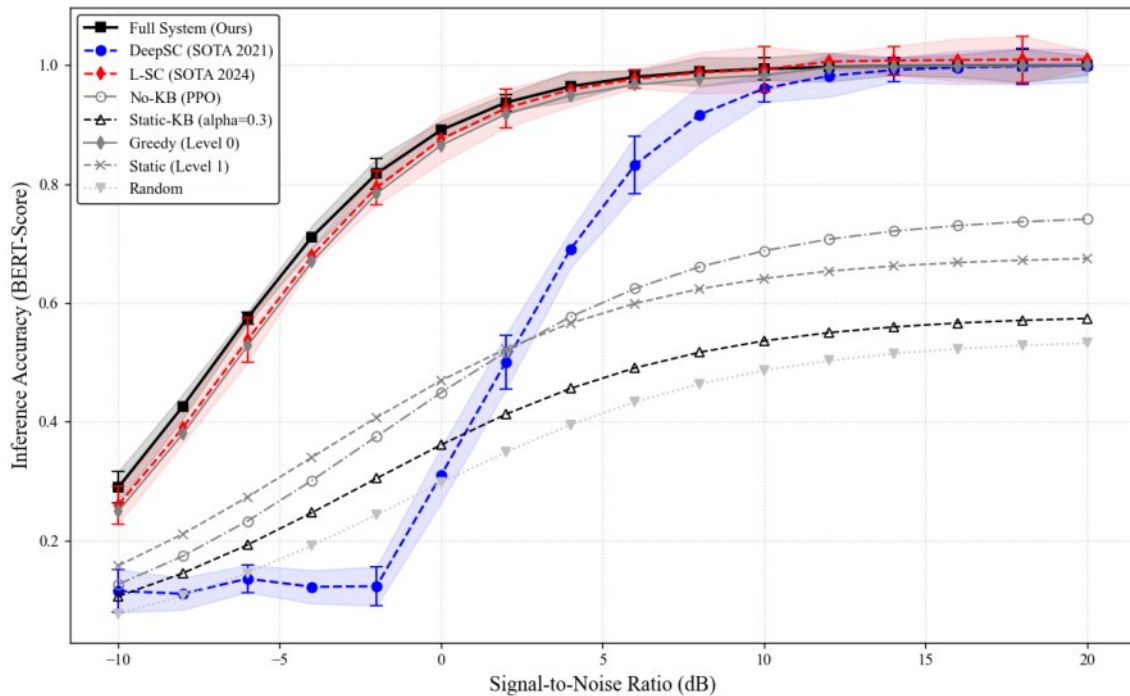


Fig. 5: SOTA Comparison under Non-stationary Channels (5 Seeds)

图5 真实视觉场景下的极端信道鲁棒性与SOTA对比实验分析



呈现出典型的对数型收敛趋势。在训练初期的前 200 个回合内，曲线斜率较大，表明智能体处于高频的探索阶段，持续将未知的高价值“信道-语义”联合物理状态通过 LSH 算子映射并沉淀至本征知识子空间；随着训练推进至 400 回合以后，曲线逐步趋于渐进稳定，形成约 15000 个独立状态的饱和平台。

这一从“高频积累”到“渐进饱和”的闭环过程，从物理系统论的层面验证了 LSH 降维算子的保距性与收敛有界性。实验表明，该动态知识库能够在严格控制内存与维度开销的前提下，高效捕获环境的底层核心规律，为处于高维非平稳环境中的边缘智能体提供稳定可靠的底层决策边界。

#### 4.3.4 实验四：高度非平稳环境下的熵驱动微观动力学

如果说前述实验侧重于宏观系统效用的验证，那么本节则深入解构智能体在面临时变约束时的微观决策更新机理。传统工程算法往往采用僵化的静态常数来融合策略与先验，本实验重点剖析了本文首创的“基于策略熵的动态融合权重”的微观自适应更新轨迹。

通过记录智能体在持续注入外部动态扰动时的内部参数变化，微观更新图谱清晰地展现了  $\alpha_t$  的非平稳波动特征。实验发现， $\alpha_t$  并未坍缩为单一常数，而是与系统内部的信息熵以及接收端的知识失配度（KL 散度）形成了高度的数理协同响应。需要强调的是，尽管  $\alpha_t$  的绝对波动区间在 0.44 至 0.58 之间，但在 Softmax 指数归一化空间中，0.14 的权重偏移足以导致动作输出概率分布发生数量级的倾斜，从而在极短的决策步长内实现探索与利用边界的瞬时反转，有效规避动作崩溃。

当物理环境发生剧烈衰落或知识对齐发生严重失配时，神经网络难以通过局部侦测输出高置信度动作，其策略输出分布的信息熵瞬间激增。此时，底层机制敏锐捕捉到这一“决策迷茫期”，自适应拉升了融合权重  $\alpha_t$ （曲线呈现显著峰值），促使智能体主动依赖本地动态语义知识库中沉淀的平稳先验经验，从而实施了边界风险规避；反之，在网络状态进入优质稳态期时，网络熵值回落， $\alpha_t$  自动衰减至低位，智能体重新释放出主动探索高阶语义表征的潜力。这一熵驱动的动态微观寻优机制，揭示了 KE-PPO 算法有效规避非平

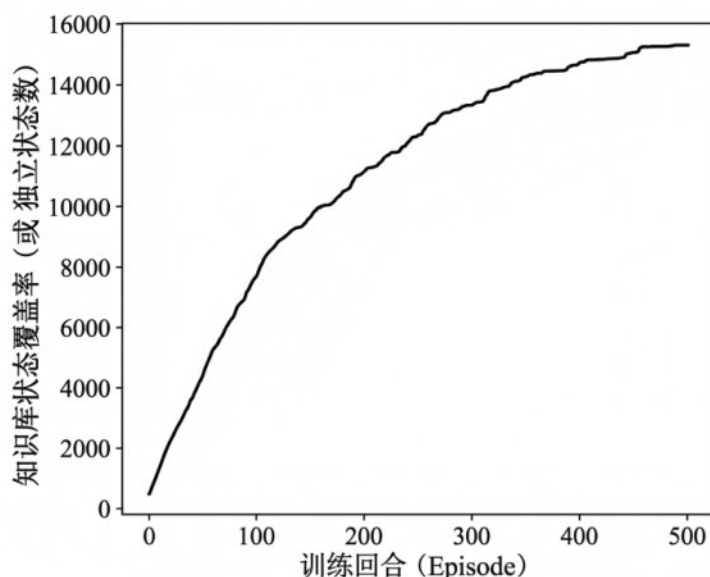


图6 知识库容量更新轨迹

稳陷阱、实现对 6G 极端动态环境深度自适应的底层逻辑。

#### 4.3.5 实验五：LSH 哈希维度 $k$ 的性能敏感性分析

为了验证将 529 维联合状态特征压缩至  $k=8$  维语义流形的合理性，本节对 LSH 降维算子的超参数  $k$  进行了消融与敏感性分析。实验分别测试了  $k \in \{4, 8, 16, 32, 64\}$  时的系统平均累计奖励与检索匹配延迟。实验结果表明，当  $k=4$  时，由于压缩过度，哈希碰撞率显著急剧上升，Critic 网络难以完全吸收局部观测噪声，导致系统累计奖励下降约 18%；当  $k=16$  时，虽然碰撞率随维度的增加而降低，系统奖励趋于饱和（仅微幅上升约 2%），但知识库的状态空间规模呈指数级膨胀，导致边缘节点的哈希检索与在线更新延迟突破了亚毫秒级约束。如下图 8，综合评估表征精度与边缘算力开销，选取  $k=8$  能够在可容忍的碰撞误差界限内，实现系统性能与检索复杂度的最佳折中。

## 5 结束语

本文针对 6G 语义通信中环境波动与边缘知识失配带来的挑战，研究了一种面向深层语义智能体的知识更新与自适应传输机制。通过引入局部敏感哈希算子确立特征降维的误差界限，并将系统决策重构为面向感知节点的单智能体 POMDP，所提 KE-PPO 算法在理论计算与仿真评估中展现出显著的信令削减潜力，其理论控制面交互量估算较传统分布式方案可实现约两个数量级的缩减。实验结果表明，在 -10 dB 极端衰落信道等特定环境下，本机制的推理准确率较 DeepSC 提升约为 16.5%，且边缘节点的在线更新算力开销基本维持在亚毫秒级。该机制在一定程度上抑制了非平稳环境下的策略震荡，在保障实时性的前提下，实现了推理精度与通信代价的有效折中。本研究从信息论视角探讨了语义传输的联合优化路径，可为未来 6G 边缘智联网络中轻量级、意图驱动节点协同提供理论参考与演

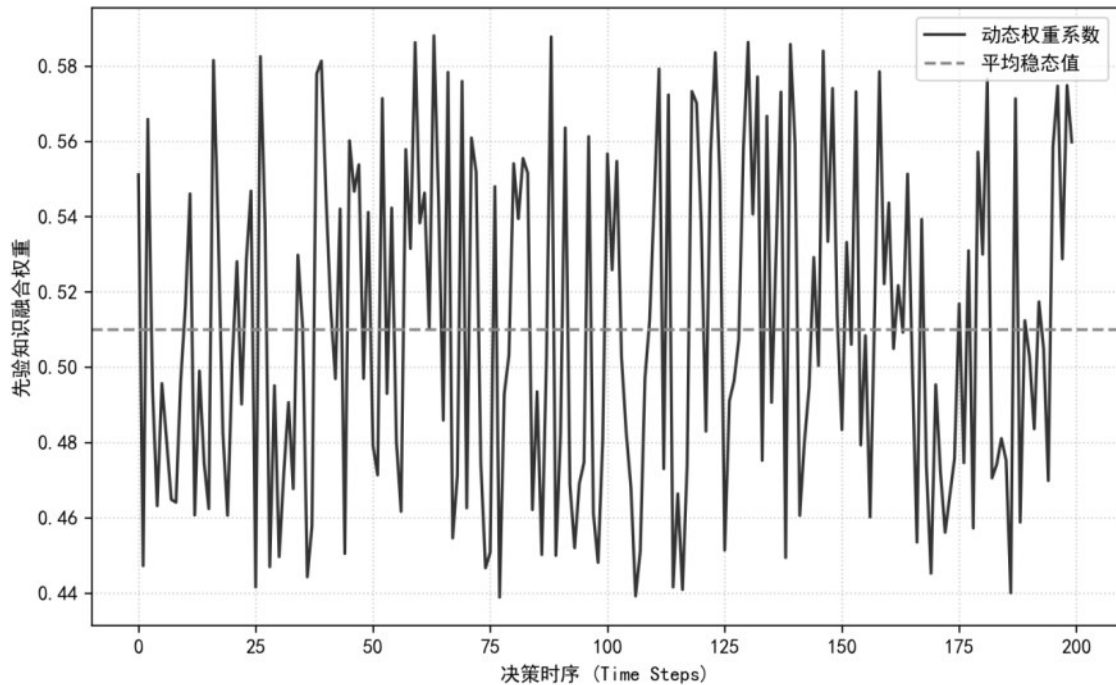


图7 高度非平稳环境下的熵驱动微观动力学

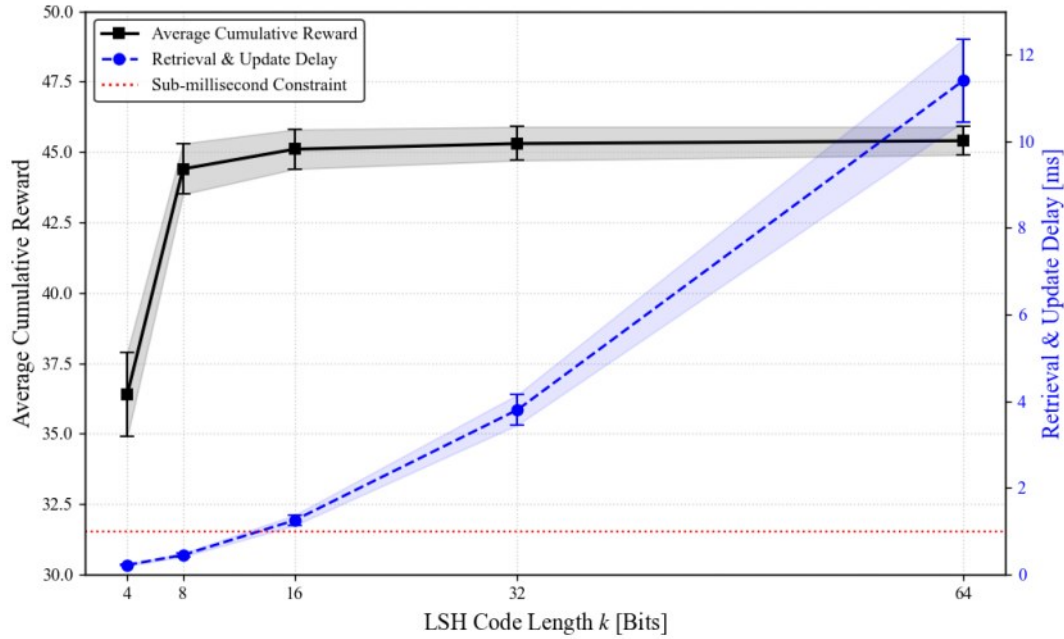


图8 双轴敏感性与计算开销折中图

进范式。

## 参考文献:

- [1] STRINATI E C, BARBAROSSA S, GONZALEZ-JIMENEZ J L, et al. 6G networks: Beyond Shannon towards semantic and goal-oriented communications[J]. Computer Networks, 2021, 190: 107930.
- [2] ZHANG P, XU W, GAO H, et al. Toward wisdom-evolutionary and primitive-concise 6G: A new paradigm of semantic communication networks[J]. Engineering, 2022, 8: 60-73.
- [3] SHAO J, MAO Y, ZHANG J. Learning task-oriented communication for edge inference: An information bottleneck approach [J]. IEEE Journal on Selected Areas in Communications, 2022, 40(1): 197-211.
- [4] GÜNDÜZ D, QIN Z, AGASOUNIS I, et al. Beyond transmitting bits: Context, semantics, and task-oriented communications [J]. IEEE Journal on Selected Areas in Communications, 2022, 41(1): 5-41.
- [5] XIE H, QIN Z, LI G Y, et al. Deep learning enabled semantic communication systems[J]. IEEE Transactions on Signal Processing, 2021, 69: 2663-2679.
- [6] YANG W, DU H, LIEW S C, et al. Semantic communications for future internet: Fundamentals, applications, and challenges [J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 213-250.
- [7] SHI G, XIAO Y, LI Y, et al. From semantic communication to se-mantic-aware networking: Model, architecture, and open problems[J]. IEEE Communications Magazine, 2021, 59(8): 44-50.
- [8] JI Y, WANG Y, ZHAO H, et al. Multi-agent reinforcement learning resources allocation method using dueling double deep Q-network in vehicular networks[J]. IEEE Transactions on Vehicular Technology, 2023, 72(10): 13447-13460.
- [9] DU H, LI Z, NIYATO D, et al. Enabling AI-Generated Content Ser-vices in Wireless Edge Networks[J]. IEEE Wireless Communications, 2024, 31(3): 226-234.
- [10] YANG Q, LUO X, et al. Joint Transmission and Understanding of Semantics with Edge Intelligence[C]// IEEE International Conference on Communications (ICC), 2023: 1-6.
- [11] JI Y, et al. Multi-Agent Reinforcement Learning for Covert Semantic Communications over Wireless Networks[C]// 2023 IEEE In-ternational Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023: 1-5.
- [12] WANG Y, CHEN M, LUO T, et al. Performance optimization for semantic communications: An attention-based reinforcement learning approach[J]. IEEE Journal on Selected Areas in Communications, 2022, 40(9): 2598-2613.
- [13] ZHOU F, LI Y, ZHANG X, et al. Cognitive semantic communication systems driven by knowledge graph[J]. IEEE Transactions on Wireless Communications, 2023, 22(3): 1234-1248.

- [14] WANG Y, et al. Semantic Communication for Partial Observation Multi-agent Reinforcement Learning[C]// 2023 IEEE Statistical Signal Processing Workshop (SSP), 2023: 1-5.
- [15] QI Q, WANG J, MA Z, et al. Knowledge-enhanced semantic communication system with OFDM transmissions[J]. Science China Information Sciences, 2023, 66(8): 182304.
- [16] YU C, VELU A, WANG B, et al. The surprising effectiveness of PPO in cooperative multi-agent games[C]// Advances in Neural Information Processing Systems (NeurIPS), 2022, 35: 24611-24624.
- [17] SHAO Y, LOURO C P, GUNDUZ D. Learning over-the-air digital semantic communication[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(8): 2490-2505.

#### [作者简介]

张子天 (1988-), 男, 博士, 浙江工商大学信息与电子工程学院 (萨塞克斯人工智能学院) 副研



究员, 主要研究方向为基于人工智能的网络流量预测与资源管理。



诸葛斌 (1976-), 男, 博士, 浙江工商大学信息与电子工程学院 (萨塞克斯人工智能学院) 教授, 主要研究方向为网络与通信技术、互联网技术和网络安全。



洪仕玉 (2002-), 女, 硕士研究生, 浙江工商大学信息与电子工程学院 (萨塞克斯人工智能学院), 主要研究方向为数据资源调度、智慧网络。



董黎刚 (1973-), 男, 博士, 浙江工商大学信息与电子工程学院 (萨塞克斯人工智能学院) 教授, IEEE 和 IEEE-CS 成员, 中国电子学会高级会员, 浙江计算机学会理事, 主要研究方向为基于大数据与深度学习的智慧网络与智慧教育及分布式系统理论。