

信息年龄敏感的稀疏移动群智感知机制^①

赵雅馨^{1,2}, 肖明军^{1,2}

¹(中国科学技术大学 计算机科学与技术学院, 合肥 230027)

²(中国科学技术大学 苏州高等研究院, 苏州 215004)

通信作者: 肖明军, E-mail: xiaomj@ustc.edu.cn



摘要: 稀疏移动群智感知是一种新兴的模式, 它从感知区域的子集收集数据, 然后推理其他区域的数据. 然而, 在实际应用中, 工人不足或分布不均的情况广泛存在. 因此, 在有限的预算下, 必须优先选择相对更重要的工人收集数据. 此外, 许多稀疏移动群智感知应用对数据的时效性要求较高. 因此本文将考虑数据的新鲜度, 并使用信息年龄作为新鲜度指标. 为了解决这些挑战, 本文提出了一种轻量级年龄敏感的数据感知和推理框架. 该框架旨在预算约束下, 选择合适的工人收集数据, 并通过准确捕捉感知数据时空关系进行数据推理, 以优化信息年龄和推理的准确性. 由于预算和工人有限, 可能会导致数据量较少的情况. 因此, 本文还提出了精简数据推理模型的方法, 以提高推理效率. 通过广泛的实验进一步论证了该框架在实际应用中的优越性.

关键词: 稀疏移动群智感知; 信息年龄; 强化学习; 工人选择

引用格式: 赵雅馨, 肖明军. 信息年龄敏感的稀疏移动群智感知机制. 计算机系统应用, 2024, 33(8): 115-122. <http://www.c-s-a.org.cn/1003-3254/9601.html>

Age-of-information-aware Sparse Mobile Crowdsensing Mechanism

ZHAO Ya-Xin^{1,2}, XIAO Ming-Jun^{1,2}

¹(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)

²(Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215004, China)

Abstract: Sparse mobile crowdsensing (MCS) is an emerging paradigm that collects data from a subset of sensing areas and then infers data from other areas. However, there is a shortage or uneven distribution of workers when sparse MCS is applied. Therefore, with a limited budget, it is important to prioritize the involvement of the more important workers in data collection. Additionally, many sparse MCS applications require timely data. Consequently, this study considers data freshness, with age of information (AoI) serving as a freshness metric. To address these challenges, a simplified AoI-aware sensing and inference (SASI) framework is proposed in this study. This framework aims to optimize AoI and inference accuracy by selecting suitable workers for data collection under budget constraints and accurately capturing spatiotemporal relationships in sensed data for inference. Moreover, limited budgets and worker availability may result in a reduced volume of data. Thus, methods for streamlining data inference models are also proposed to enhance inference efficiency. Experiments have substantiated the superiority of this framework in practice.

Key words: sparse mobile crowdsensing (MCS); age of information (AoI); reinforcement learning; worker selection

1 引言

移动群智感知 (mobile crowdsensing, MCS)^[1]是一

种经济高效的大规模数据采集方法, 已被应用于环境监测^[2]、交通管理^[3]、城市感知等多个领域. 通常来说,

① 基金项目: 国家自然科学基金面上项目 (62172386, 61872330); 江苏省自然科学基金面上项目 (BK20231212)

收稿时间: 2024-02-25; 修改时间: 2024-03-28; 采用时间: 2024-04-10; csa 在线出版时间: 2024-07-03

CNKI 网络首发时间: 2024-07-08

人,工人在其所在的区域 n_i 进行数据采集.

数据收集: 在第 2 阶段,工人收集数据并上传到平台.平台将存储接收到的数据,并更新感知区域对应的 AoI 值(在第 1.2 节中定义).之后,工人将收到佣金.在图 1 的示例中,工人对应的佣金等于 c_1 . n_1 的 AoI 值从 8 下降到 0,而其他感知区域的 AoI 值则随着时间的推移呈线性增长.

数据推理: 我们使用数据推理模型来推理未采集的数据.在图 1 的示例中,除了左上方的感知区域是最新收集的数据外,其余区域的数据都是推理出来的.然后,将推理结果发送给服务请求者.

模型更新: 最后,推理模型利用收集到的数据进行实时更新.经过一轮轮的更新,模型变得更加精确.

2.2 AoI 模型

为了衡量数据的新鲜度,我们将 $\Delta_i(t)$ 定义为区域 n_i 的瞬时 AoI,它可以通过最近一次更新后的时间来计算.更新策略指明了每次由工人上传哪个感知区域的数据.我们将 $\pi_i(t)$ 定义为二进制变量,表示感知区域 n_i 的数据是否在时间 t 上传,而 $\pi(t)=\{0, 1, 0, \dots, 0\}$ 表示所有感知区域的更新策略.为便于解释,我们假定 $\Delta_i(t)$ 在更新发生后下降为 0,即时间为 t . $\Delta_i(t_m)$ 的演变过程可描述为:

$$\Delta_i(t_m) = \begin{cases} t_m - t_{m-1} + \Delta_i(t_{m-1}), & \text{if } \pi_i(t_m) = 0 \\ 0, & \text{if } \pi_i(t_m) = 1 \end{cases} \quad (1)$$

其中, t_m 是数据在第 m 轮 ($m \geq 1$) 上传到平台的时间.特别地,当 $m=1$ 时, $t_{m-1}=0$, $\Delta_i(t_{m-1})=0$.此外 $\pi_i(t_m)$ 表示在 t_m 时是否选择收集区域 n_i 的数据.

2.3 问题定义

在上述稀疏 MCS 系统中,平台逐轮收集数据.给定预算限制 B 和持续时间 T ,我们的目标是在线选择一组工人在目标感知区域收集数据,使得感知数据的平均 AoI 和数据推理损失最小化,形式化如下:

$$\begin{cases} \min \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \Delta_i(t_m) + \sum_{t=1}^T \beta L(t_m) \\ \text{s.t.} \begin{cases} \hat{M} = f(M) \\ w_j \in W, \sum_{w_j=1}^k \pi_j c_j \leq B \end{cases} \end{cases} \quad (2)$$

其中, β 是为平衡平均 AoI 和推理损失之间的权重因子.这里, M 表示存储在平台中的原始数据,而 \hat{M} 表示

通过数据推理方法 $f(\cdot)$ 获得的推理数据.推理损失用 L 表示,它量化了时间 t 推理数据 \hat{M} 与实际采集数据 M 之间的差异. W 是在时间 T 内连续到达的工人集合, w_i 是第 i 个到达的工人.此外, $|\pi|$ 表示到当前为止所选工人的累积计数, $\pi_j \in \{0, 1\}$ 代表是否选择 j 个工人. $\Delta_i(t)$ 遵循式 (1) 中指定的公式.

3 在线年龄敏感的数据感知和推理框架

为了解决上述问题,我们提出轻量级年龄敏感的数据感知和推理 (simplified AoI-aware sensing and inference, SASI) 框架,该框架的设计旨在实现轻量且高效的推理模型,同时考虑数据的新鲜度作为感知决策和数据推理的重要因素.具体而言,该框架主要以强化学习算法深度 Q-learning 网络 (DQN) 为基础.我们将框架抽象为一个强化学习 (RL) 代理,其根据已有数据 (状态) 选择一个行动 (工人),以最小化 AoI 和推理损失 (奖励).因此,RL 代理能够招募工人收集数据,并根据环境反馈学习有效的策略,以及根据当前状态预测未来.总体而言,SASI 包括两个模块:工人选择和数据推理.

一方面,工人选择是由代理的策略网络结合秘书算法决定的.具体来说,代理的策略网络本质上是从状态空间到行动空间的函数映射.鉴于在我们的场景中,行动空间对应于出现在某个感知区域的工人,因此可以使用分类器来表示.这种情况下,我们使用 Transformer 神经网络来表示代理的策略网络,它可以考虑不同区域之间的全局关系,再结合秘书算法对于每个到来的工人做出决策.更多详情参见第 2.1 节.

我们还将数据推理模块设置为 Transformer 神经网络.它将当前状态下所有感知区域的数据作为输入,并在下一个时间步返回所有区域的推理结果.我们在输入中嵌入了位置信息和 AoI,因此神经网络可以有效捕捉时空信息,并且针对稀疏 MCS 的小数据量场景,我们改进了精简的模型,从而实现精确推理.更多细节将在第 2.2 节中详细阐述.

在代理与环境的交互过程中,代理利用数据推理模型对所有感知区域进行推理.基于工人选择的策略网络,它决定实际招募哪个区域的工人收集数据,该区域称为区域 X .然后,代理将收集到的数据与 X 区域推理数据之间的误差以及所有区域的 AoI 结合起来,形成奖励.该奖励用于优化策略网络和数据推理模型,使代理能够改进有效的策略.

3.1 工人选择

我们要解决的第1个难题是选择哪个工人收集数据,以获得最大的未来奖励。首先,对于每个即将到来的工人,我们首先从历史记录和即将到来的工人中进行一次抽样。然后,我们利用强化学习根据工人对数据推理的贡献估计值和先知秘书策略选择合适的工人。

3.1.1 模型定义

不同工人所在的感知区域的数据对于数据推理的具有不同的贡献度。例如,来自AoI较大区域的信息往往相对过时,对提高数据推理的准确性贡献较小。因此,应优先考虑选择AoI较高区域的工人。由于数据的空间相关性,最好招募位置分散的工人收集数据,以提高信息量。我们使用强化学习来解决这个问题,因为它不仅考虑了过去的信息,还考虑了未来的奖励。

我们将“基于AoI的工人贡献度估计”问题抽象为一个RL代理,该代理根据已收集的数据(状态)选择一个区域(行动),以最小化AoI和推理损失(奖励)。

状态:表示当前采集的数据和相应的AoI值。状态空间用矩阵 $S_{m \times 2}$ 表示,每行对应一个感知区域,每列代表状态空间的一个元素。其中第1列是最新采集的若干感知区域数据 y_i ;第2列是区域 i 数据的AoI A_i 。

因此,状态空间可以表示为:

$$S = [y, A] \quad (3)$$

行动:指的是决定在当前一轮中选择哪个工人进行数据采集。因此,行动空间可以定义为所有可能的工人所在感知区域的集合 $\mathbb{A} = \{1, 2, \dots, m\}$ 。

$$R = -\frac{1}{N} \sum_{i=1}^N A_i(t) - \beta L(t) \quad (4)$$

奖励:用来表示一个行动的好坏。由于我们的目标是 minimized 平均AoI和推理损失,因此可以将奖励定义为式(2)第1个子式的负值,如下所示。

代理的目标是学习一种策略,将每个状态映射为最佳行动,以最大化预期累积奖励。该策略由一个深度神经网络表示,该网络将状态作为输入,并输出行动空间的概率分布。代理根据该分布选择概率最高的行动,采用探索-利用策略,在利用当前最佳行动和探索新行动之间取得平衡,从而不断改进策略。

3.1.2 秘书算法

在在线场景中,工人不断到来,我们需要实时决定选择或者不选择一个工人。因为我们有预算限制 B ,因

此可以将问题抽象为从连续到来的 Y 个工人中选择 k 个,使得对数据推理的贡献度最高,且花费满足式(2)第3个子式。我们将工人 w_i 的贡献定义如下:

$$f(w_i) = R \quad (5)$$

其中, R 是基于第2节的定义模型的行动获得的奖励。

传统的秘书问题从 y 个工人里选出一个最佳工人,解决策略是:对于某些整数 r ,其中 $1 \leq r \leq y$ 。先面试前 r 人,都不聘请他们,在之后的 $n-r$ 人中,如果任何一人比前面的人都更佳,便聘请他。在我们的情景中,我们将这一概念扩展到选择 k 个工人^[10]。因此,我们将 Y 个工人分为 Y/k 段,并使用秘书问题算法从每段中选择最佳的一个工人。此外,我们引入先知秘书问题思想^[11],构造样本工人集,利用工人历史记录学习一个工人贡献分布。这种方法的优点在于,我们不再直接丢弃每段的前 r 个工人,而是通过历史数据集中的抽样来做出决策,从而提高数据利用率。

3.2 数据推理

SASI的推理模块是基于Transformer实现的。它从当前状态下的所有感知区域获取输入数据,并在随后的时间步中对所有区域进行推理。值得注意的是,由于在输入中嵌入了位置信息和AoI,它能有效捕捉时空信息,并且Transformer善于捕捉长期关系,这些是我们选择它做推理模型的理由。更重要的是,稀疏MCS场景下的数据量相对于训练有着复杂结构和大量参数的Transformer模型来说是不足的,训练不充分可能会带来过拟合的问题,因此本文中我们对传统Transformer模型做了大量简化,使得在小数据量场景下能训练出更加高效准确的模型。

3.2.1 整体结构和改进思路

SASI是基于传统Transformer的轻量级数据推理模型。它由3大块组成:嵌入块、编码器块和推理块。如图2所示,嵌入块由位置嵌入、AoI嵌入和值嵌入3部分信息组成。编码器块由两个基本块组成,每个基本块包含一个多头注意力层和一个前馈层。推理块包含了一个线性层和Softmax函数。图2中从左到右是数据流的流向。我们将原始数据通过嵌入块获得时空编码特征,然后将其送入编码器块学习数据间时空关系,最后送入推理块推理下一次应收集的感知区域的数据值。

为了应对稀疏数据场景下的复杂模型过拟合,我们针对图2中的编码器块中的基本块进行简化。传统

的 Transformer 一般遵循一个设计规范,即首先构造一个基本块,这个基本块通常由注意力层、MLP 层、残差连接和归一化层等组件构成,它们以特定方式进行精确的排列组合,随后对基本块进行堆叠形成最终的 Transformer 模型,如图 3(a) 所示.这种复杂性导致了架构的脆弱性,看似微小的变化都会大大降低训练速度,或导致模型无法训练.而结合信号传播理论和广泛的消融实验的最新研究,我们发现可以在不降低训练速度和模型表现的情况下移除 Transformer 许多块组件,包括残差连接、投影或值向量参数、顺序子块和归一化层.并且,深度缩放研究^[12]表明,当模型深度增加时,性能得到提升.说明简化的模块不仅训练速度更快,而且能有效利用更深层次的额外容量.

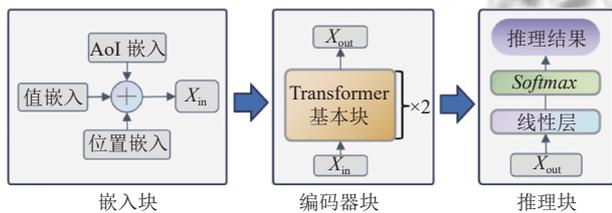


图 2 AoI 敏感的推理模型

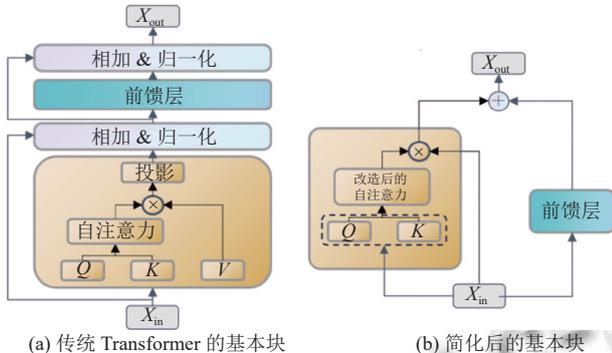


图 3 推理模型基本块简化前后对比

除此之外,我们没有采用经典的 Transformer 编码器-解码器架构,原因有两个.首先,我们的推理任务需要固定长度的输出,这可以由编码器有效处理.因此,在这种情况下不需要使用解码器来处理非固定长度的序列到序列输出.其次,在数据规模相对较小的稀疏 MCS 场景中,参数量较多但数据量有限的解码器可能会导致过拟合,降低训练效率.因此,我们利用 Transformer 编码器直接将输入序列编码为固定长度的特征向量,然后将其输入多层感知器进行回归预测.

3.2.2 嵌入块

在稀疏 MCS 中,离散或连续数据存在于不同的向

量空间.因此,我们采用嵌入技术将信息映射到相同的矢量空间,将其表示为固定长度、相同的数量级的低维连续矢量.我们将 3 种类型的信息嵌入到编码器输入中:二维位置嵌入、AoI 嵌入和值嵌入.

在值嵌入方面,我们采用线性神经网络将感知任务中的离散或连续数值映射到 d 维向量中.至于二维位置嵌入,我们分别嵌入每个感知区域的经度和纬度.在 AoI 嵌入中,我们将每个感知区域的 AoI 映射为一个 d 维向量.然后,我们将 4 个向量汇总,得到感知区域的最终嵌入向量 $X_{in} \in \mathbb{R}^{m \times d}$,其中 m 是感知区域的数量, d 是每个感知区域的向量维度.

3.2.3 简化的编码器块和推理块

编码器模块由 3 个基本块组成,每个基本块包括两层:多头注意力层和前馈层.接下来介绍对基本块的简化,包括移除残差连接、投影和值向量参数、顺序子块和归一化层,如图 3(b) 所示.

首先,自注意力的残差连接可以被移除.矩阵的秩反映了流经网络的信号的维数.从数学的角度解释,如果注意力矩阵的秩显著降低,则意味着该矩阵更接近于低维子空间,发生了秩坍塌^[13],从而失去了捕捉稀疏 MCS 场景里复杂的数据时空相关性的能力.先前有工作已经证明,直接移除自注意力的残差连接会导致模型训练发生秩坍塌现象^[13],导致可训练性变差^[14],因此需要对自注意力进行修改. Transformer 原有的注意力机制可以使用下面的矩阵形式表现:

$$A(X) = \text{Softmax} \left(\frac{XW^Q(XW^K)^T}{\sqrt{d_k}} \right) XW^V \quad (6)$$

其中, W^Q 、 W^K 、 W^V 分别为查询、键和值的权值矩阵, d_k 为键的维数,用于缩放 Softmax 函数中的点积.使用 Value-SkipInit^[15]来修改自注意力矩阵来缓解信号退化:

$$A'(X) = (\alpha I_T + \beta A(X)) \quad (7)$$

可训练标量 α 和 β 分别初始化为 1 和 0, I_T 是单位矩阵,这里的关键在于通过初始化自注意力矩阵,鼓励输入信息相对于其他信息更加关注于自身的变化,从而在更深的网络前向传播时实现更好的信号传递.这种策略只能在网络初始化阶段使用.

但是,无法忽略的一点是,移除注意力子块中的残差连接会直接导致网络整体的训练速度下降.根据研究^[15]表明,如果将更新参数限制在初始化的投影和值向量上,就可以更好的恢复跳连的效果.

第二, 将投影和值向量参数移除. 如果去掉数值和投影参数, 会简化 Transformer 的内部结构. 将这些矩阵设置为恒等式本质上将 Transformer 的相关组件转换为线性层, 从而降低了它们需要执行的操作的复杂性. 通过将值矩阵 W^V 和投影矩阵 W^P 设置为单位矩阵, 大大简化了自注意力机制. 新的公式变成:

$$Atten_{identify}(X) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)X \quad (8)$$

其中, Q 和 K 是查询矩阵和键矩阵, 通常计算为 $Q=XW^Q$ 和 $K=XW^K$. 在这种情况下, 不使用值矩阵 V , 而是使用输入矩阵 X 本身. 将 $Softmax$ 函数应用于 Q 和 K 的标量点积, 并将结果乘以 X . 同样的, 在计算多头注意力矩阵时也不再使用投影矩阵 P . 在这里, 移除投影和值向量参数, 使计算变成线性操作, 降低了注意力机制的复杂性. 从方程中消除 W^V 和 W^P 减少了大约 13% 的可训练参数数量, 从而获得更快的收敛速度.

第三, 将 MLP 子块的残差连接移除. 对于 MLP 层的残差连接优化, 一个很直接的做法就是采用并行子块设计, 同时计算 MHA 和 MLP 层的输出结果, 随后对它们进行加权求和, 这种方式在一些大型 Transformer 模型中以及得到验证, 从数学形式上定义, 并行 MLP 层的计算过程如下:

$$X_{out} = \alpha_{comb}X_{in} + \beta_{FF}MLP(Norm(X_{in})) + \beta_{SA}MHA(Norm(X_{in})) \quad (9)$$

其中, α_{comb} 、 β_{FF} 和 β_{SA} 是可训练的参数, 分别控制输入、 MLP 和 MHA 块对输出的贡献比例. $Norm$ 表示在 MLP 和 MHA 块处理之前应用于输入的归一化函数. MLP 和 MHA 块并行处理相同的归一化输入, 它们的输出与缩放后的输入求和后得到 X_{out} . 默认情况下, 残差连接权重 $\alpha_{comb}=1$, MLP 和 MHA 权重 $\beta_{FF}=\beta_{SA}=1$. 与按顺序计算子块的标准基础块相比, 通过并行处理 MHA 和 MLP 块, 训练速度提高了 15%.

第四, 将归一化层移除. 在基础块每个子块中都会使用归一化层作为后处理, 如果能去除归一化层, 就可以得到最简单的标准块. 从信号传播初始化的角度来看, 归一化操作可以隐含的削减上一子块 (残差分支) 的权重, 而这种效果也可以在残差连接过程通过明确降低残差分支的权重来实现, 或者使用前面改造后的线性自注意力来代替. 由于在前面修改过程中考虑到了这些机制 (如降低 MLP 的权重 β_{FF} 和使用改造后的线性自注意力), 因此无需再进行归一化处理.

4 实验分析

4.1 实验描述

为了评估 SASI 框架的性能, 我们在真实世界的数据集上进行了综合评估, 比如北京 36 个固定站点收集的空气质量数据^[16], 包括了 $PM_{2.5}$ 、 CO 、 SO_2 和 NO_2 的测量值. 我们选择了该数据集的一个子集, 其中包含 169 950 条记录, 代表 2013 年 2 月 8 日–2014 年 2 月 8 日期间每天和每小时记录的各空气质量参数的测量值. 每个数据条目包括站点 ID、采集时间戳和 4 个空气质量参数的值. 此外, 还有一个站点数据集, 提供了站点 ID、站点名称、纬度和经度.

上述数据集最初是从静态监测站获得的. 实际上, 这些数据也可以通过手机收集. 因此, 在实验中, 我们提出了一个合理的假设, 即稀疏 MCS 的工人可以利用其移动设备中的传感器来收集空气质量和交通数据.

4.2 对比算法

为全面了解数据推理和工人选择之间的相互作用, 我们首先对数据推理部分进行了单独实验, 然后在数据推理策略固定的情况下对工人选择部分进行了实验.

在数据推理模块, 我们选择了 5 种算法作比较方法. KNN-T 和 KNN-S 利用从 k 条时间和空间最近的数据记录中获取的数据的加权平均值, k 设置为 5. LR 是一种统计模型, 模拟时空和数据间的关系, 学习率设定为 0.01. 时空矩阵补全算法 (spatio-temporal matrix completion, STMC)^[8] 利用数据时空关系, 使用数学方法对数据进行补全. U-SAI (unabridged SAI) 是 SASI 中数据推理未简化版本模型的算法简称, 推理模型采用了传统的 Transformer, 其基本块如图 3(a) 所示.

在工人选择模块, 我们固定数据推理策略为简化的 Transformer 模型, 使用 4 种工人选择算法作比较方法. 随机策略 (RAND) 随机选择工人, 在空间维度能提供有价值的信息. 传统秘书策略 (TSS)^[10] 是不带有先知数据池、利用秘书问题算法来动态选择最佳解决方案的方法. 近最优线下策略 (NOS) 是一种在离线场景中接近最佳性能的方法.

4.3 实验结果

4.3.1 数据推理

我们评估了算法在 4 个数据集的数据推理损失, 如图 4 所示. 从推理损失来看, 我们的 SASI 算法获得了最低的推理损失. 对比算法中损失从低到高依次是

U-SAI、STMC、LR 和 KNN 算法. 由于 SASI 和 U-SAI 的深度学习模型能更好地抓住数据的时空关系, 因此优于 STMC、LR 和 KNN 算法; 此外, SASI 相比于 U-SAI 参数量减少了 13%, 更适合稀疏 MCS 的小规模数据, 表现出了比 U-SAI 更好的推理结果.

4.3.2 工人选择

我们从奖励的角度评估了工人选择算法的性能, 奖励由 AoI 和推理损失组成. 评估结果如图 5 所示. 我

们设置平衡系数为 0.02, 工人平均成本为 2, 预算约束 B 为 400, 评估了在 4 个不同数据集上的效果. 可以看出, 最初由于冷启动阶段的数据量有限, 所有算法的奖励都很差. 随着收集数据和在线更新, 所有算法性能都有所提高. 其中, NOS 是拥有完全信息的理想情况, 因此奖励最高. 除了 NOS 之外, 我们的 SASI 由于其基于历史信息能在在线场景中做出合适决策的优势, 在所有数据集上获得了更优的奖励.

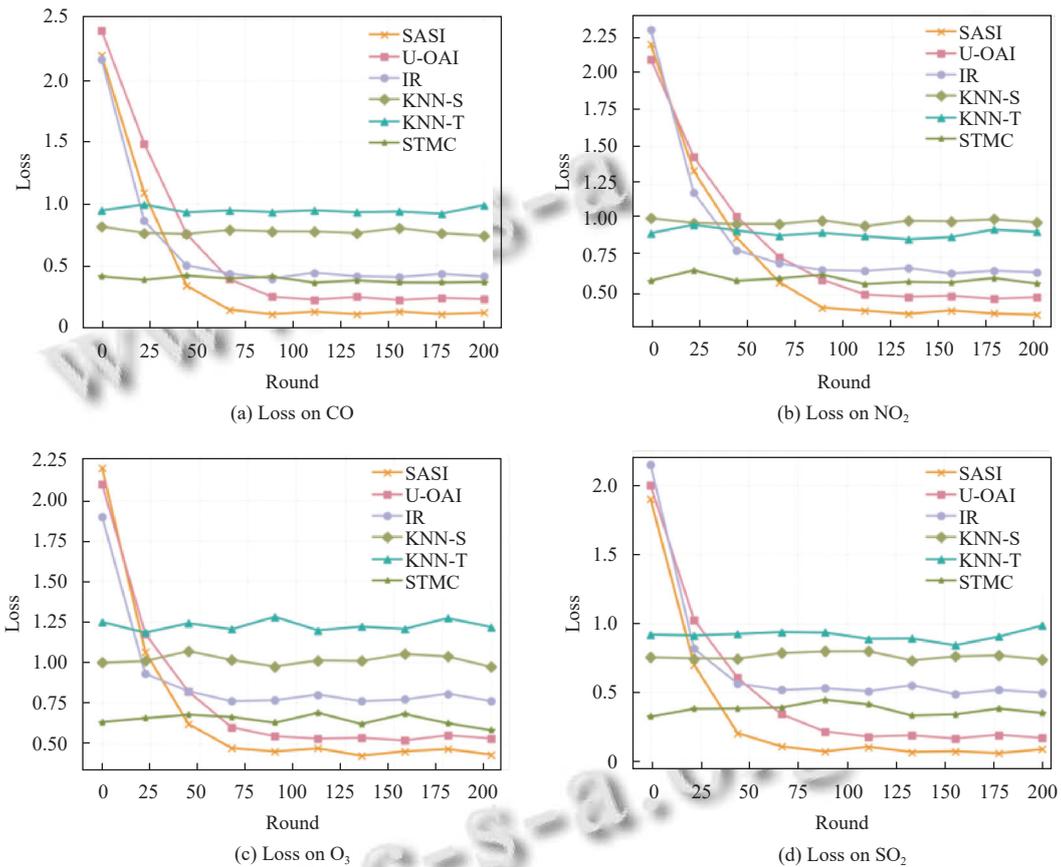


图 4 不同数据推理算法在不同数据集上的损失

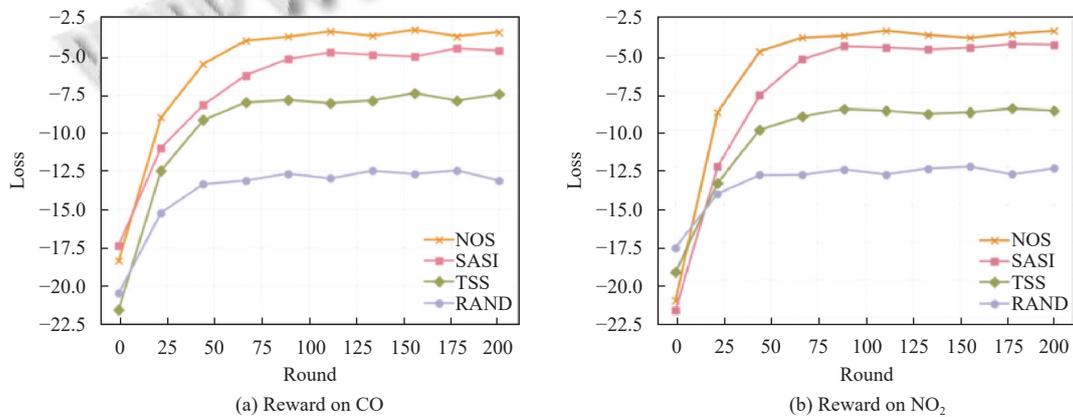


图 5 不同工人选择算法在不同数据集上的回报

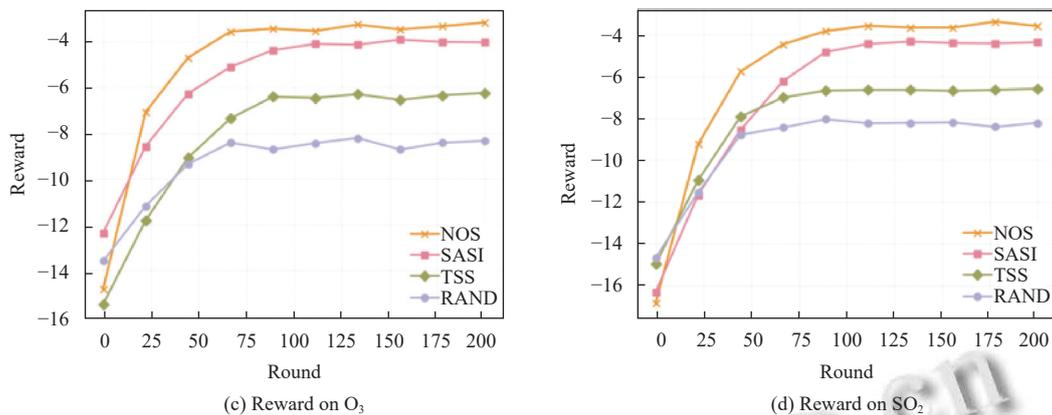


图5 不同工人选择算法在不同数据集上的回报(续)

5 结论与展望

本文针对稀疏 MCS 中的工人选择和数据新鲜度展开研究,提出了轻量级年龄敏感的数据感知和推理(SASI)框架。SASI 框架通过合适的工人选择和数据推理优化信息年龄和推理准确性,应对了预算约束和工人不足的挑战,并通过模型精简应对小数据量的问题。实验结果表明,SASI 在性能上优于几种基准算法。

参考文献

- Capponi A, Fiandrino C, Kantarci B, *et al.* A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities. *IEEE Communications Surveys & Tutorials*, 2019, 21(3): 2419–2465.
- Wang E, Liu WT, Liu WB, *et al.* Spatiotemporal urban inference and prediction in sparse mobile crowdsensing: A graph neural network approach. *IEEE Transactions on Mobile Computing*, 2023, 22(11): 6784–6799.
- Xue JZ, Xu YT, Wu W, *et al.* Sparse mobile crowdsensing for cost-effective traffic state estimation with spatio-temporal transformer graph neural network. *IEEE Internet of Things Journal*, 2024, 11(9): 16227–16242. [doi: [10.1109/JIOT.2024.3356554](https://doi.org/10.1109/JIOT.2024.3356554)]
- Zhao ST, Qi GZ, He TJ, *et al.* A survey of sparse mobile crowdsensing: Developments and opportunities. *IEEE Open Journal of the Computer Society*, 2022, 3: 73–85. [doi: [10.1109/OJCS.2022.3177290](https://doi.org/10.1109/OJCS.2022.3177290)]
- Kaul S, Yates R, Gruteser M. Real-time status: How often should one update? *Proceedings of the 2012 IEEE INFOCOM*. Orlando: IEEE, 2012. 2731–2735.
- Yates RD, Sun Y, Brown DR, *et al.* Age of information: An introduction and survey. *IEEE Journal on Selected Areas in Communications*, 2021, 39(5): 1183–1210. [doi: [10.1109/JSAC.2021.3065072](https://doi.org/10.1109/JSAC.2021.3065072)]
- Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- Wang LY, Zhang DQ, Yang DQ, *et al.* SPACE-TA: Cost-effective task allocation exploiting intradata and interdata correlations in sparse crowdsensing. *ACM Transactions on Intelligent Systems and Technology*, 2018, 9(2): 20.
- Xie K, Tian JZ, Xie GG, *et al.* Low cost sparse network monitoring based on block matrix completion. *IEEE Conference on Computer Communications*. Vancouver: IEEE, 2021. 1–10.
- Liu WB, Yang YJ, Wang E, *et al.* Dynamic online user recruitment with (non-) submodular utility in mobile crowdsensing. *IEEE/ACM Transactions on Networking*, 2021, 29(5): 2156–2169. [doi: [10.1109/TNET.2021.3083955](https://doi.org/10.1109/TNET.2021.3083955)]
- Ehsani S, Hajiaghayi M, Kesselheim T, *et al.* Prophet secretary for combinatorial auctions and matroids. *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms*. New Orleans: SIAM, 2018. 700–714.
- He B, Hofmann T. Simplifying transformer blocks. *arXiv:2311.01906*, 2023.
- Dong YH, Cordonnier JB, Loukas A. Attention is not all you need: Pure attention loses rank doubly exponentially with depth. *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021. 2793–2803.
- Noci L, Anagnostidis S, Biggio L, *et al.* Signal propagation in Transformers: Theoretical perspectives and the role of rank collapse. *Proceedings of the 36th Conference on Neural Information Processing Systems*. New Orleans, 2022.
- He B, Martens J, Zhang GD, *et al.* Deep Transformers without shortcuts: Modifying self-attention for faithful signal propagation. *Proceedings of the 11th International Conference on Learning Representations*. Kigali: OpenReview.net, 2023.
- Zheng Y, Liu FR, Hsieh HP. U-air: When urban air quality inference meets big data. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Chicago: ACM, 2013. 1436–1444.

(校对责编: 孙君艳)