arXiv:2305.04083v2 [cs.IT] 9 May 2023

目标导向张量：超越信息年龄，迈向语义赋能的目标导向通信

Aimin Li、Shaohua Wu 和 Sumei Sun

摘要——**信息源动力学、不可靠信道与信息陈旧性之间错综复杂的相互作用，长期以来一直被认为是接收端实现准确、及时且最重要的是目标导向决策的一大障碍。因此，涌现出许多颇具前景的指标，例如信息年龄、信息价值和均方误差，用于量化这些潜在的不利因素。沿着这一思路，对这些指标进行优化间接提升了目标导向决策的效用。然而，迄今尚无专门设计用于直接评估目标导向决策过程效用的指标。为此，本文研究了一种新型性能指标——目标导向张量（GoT），用于直接量化语义失配对目标导向决策的影响。基于 GoT，我们考虑一对采样器—决策者，二者协同且分布式地工作，以实现通信的共同目标。我们构建了一个无限时域分散式部分可观测马尔可夫决策过程（Dec-POMDP），以联合推导最优确定性采样策略和决策策略。仿真结果表明，在目标达成效用和稀疏采样率两方面，采样器—决策者协同设计均优于现有关于信息年龄及其变体的研究，这标志着稀疏采样器与目标导向决策者协同设计取得了显著成果。**

索引词—**目标导向通信、目标导向张量、状态更新、信息年龄、错误信息年龄、信息价值。**

I. 引言

信息年龄（AoI）是 [1], 中提出的一项指标，现已成为衡量接收端感知到的数据新鲜度的重要指标。自提出以来，信息年龄受到了广泛关注，并被广泛分析和优化，以提升排队系统、物理层通信、MAC 层通信、物联网等的性能 [2]。这些研究工作的共同出发点是：新近接收到的消息通常包含关键且有价值的信息，从而提高决策过程的准确性和及时性。

尽管信息年龄已被证明在许多对新鲜度要求较高的应用中十分有效，但它仍存在若干关键缺陷。具体而言，(*a*)信息年龄无法直接衡量信息价值；(*b*)信息年龄未考虑源数据的内容动力学；(*c*)信息年龄忽略了端到端（E2E）信息不匹配对决策过程的影响。为解决这些局限性，研究人员对信息年龄的众多变体进行了广泛研究。其中一种典型方法是对信息年龄施加非线性惩罚 [3]–[5] 以评估陈旧信息所导致的端到端“不满”程度。这种非线性惩罚称为信息价值（VoI），有助于缓解上述缺陷（a）。其他研究则尝试解决缺陷（b） [6], [7]。在 [6], 中，研究人员提出了信息变化年龄（AoCI），以解决信息年龄忽略内容动力学的问题。在这一指标下，未发生变化的状态不一定能提供新信息，因此不会被优先传输。在 [7], 中，作者提出了一种新型年龄惩罚指标同步年龄（AoS），用于表示自最近一次同步后经过的时长。均方误差（MSE）及其变体被引入，以解决缺陷（c） [8], [9]。在 [8], 中，作者引入上下文感知的加权系数，提出信息紧迫性（UoI），用于衡量不同紧迫性水平上下文下的加权 MSE。此外，考虑到端到端不匹配可能会随着时间推移对整个系统的性能产生不利影响， [9]的作者提出了一种名为错误信息年龄（AoII）的新指标，用于量化端到端不匹配持续时间所造成的负面影响。AoII指标表明，端到端语义不匹配的程度和持续时间都会导致后续决策中的效用降低。

尽管取得了上述进展，端到端不匹配如何影响决策效用这一问题仍未得到探讨。 [10]–[13] 提出了一项名为错误执行成本 的指标，以深入研究因实时估计不精确而导致的错误执行成本。具体而言，错误执行成本 由一个非对称零对角矩阵$\mathbf{C}$表示，其中每个值$C_{X_t,\ \hat{}X_t}$表示端到端不匹配状态（X_t，$\hat{}X_t$）下的瞬时成本$X_t \neq \hat{}X_t$。对此，作者揭示，决策效用 与端到端语义不匹配类别密切相关，而非不匹配持续时间（信息不准确年龄）或不匹配持续时间（MSE）。例如，将“有火灾”估计为“无火灾”这一端到端语义不匹配类别会导致较高成本；相反的情形则会导致较低成本。

尽管如此，我们注意到： i) 如何求得错误执行成本仍不明确，这隐含地要求预先制定决策策略； ii) 错误执行成本 未考虑上下文变化因素，而这些因素也可能影响决策效用； iii) 矩阵的对角线为零这一性质意味着如下假设：如果

$\hat{X}_t = X_t$，那么 $C_{X_t, \hat{X}_t} = 0$，这意味着无误差执行不需要消耗能量。为了解决这些问题，本文作者最近在 [14]，中提出了一种名为 GoT 的新指标。与错误执行成本相比，该指标引入了上下文 Φ_t 和决策策略 π_A 这两个新维度，用以描述决策的真实效用 。本文进一步从技术层面探究 GoT 的潜力，主要内容如下：

- 我们采用 GoT，重点研究决策效用 问题。我们观测一个受控马尔可夫源，其状态转移不仅取决于接收端的决策，还取决于其所处的上下文。在这种情况下，决策会从三个方面影响效用： *i*) 信源的未来演化； *ii*) 信源的瞬时成本； *iii*) 执行所消耗的能量和资源。

- 我们实现了面向目标的采样器—决策者协同设计；据我们所知，这是首项研究采样与决策之间权衡的工作。具体而言，我们将该问题建模为一个双智能体无限时域分散式部分可观测马尔可夫决策过程，其中一个智能体代表采样器，另一个代表决策者。值得注意的是，即使对于有限时域的分散式部分可观测马尔可夫决策过程，求得最优解也已知是 NEXP 完全 [15],；我们设计了一种 RVI 暴力搜索算法，以获取最优的联合采样与决策策略。

II. 系统模型

我们考虑一个时隙化的感知—执行回路，其中感知到的语义 $X_t \in \mathcal{S} = \{s_1, \cdots, s_{|\mathcal{S}|}\}$ 和上下文 $\Phi_t \in \mathcal{V} = \{v_1, \cdots, v_{|\mathcal{V}|}\}$ 均输入语义采样器，由其判断当前状态 X_t 的重要性，并据此决定是否通过不可靠信道进行传输。语义和上下文均被提取出来，并假定能够准确描述所观测过程的状态。二元指示变量 $a_S(t) = \pi_S(X_t, \Phi_t, X_t) \in \{0, 1\}$表示时隙 t中的采样和传输动作，其中值为 1 表示执行采样和传输，值为 0 表示采样器处于空闲状态。 π_S 表示采样策略。我们考虑一个无延迟的理想反馈信道 [10]–[13]，其中 ACK 表示成功传输，NACK 则表示传输未成功。接收端的决策者将基于估计值 $\hat{X}_t$做出决策$a_A(t) \in \mathcal{A}_A = \{a_1, \cdots, a_{|\mathcal{A}_A|}\}$，这最终会影响系统的效用 。图1展示了所考虑模型的示意图。

A. 语义与上下文动态

我们考虑一个受控的离散马尔可夫源：

$$\Pr\left(X_{t+1} = s_u \,|X_t = s_i, a_A(t) = a_m, \Phi_t = v_k\right) = p_{i,u}^{(k,m)}. \tag{1}$$

此处，源的动力学取决于决策 $a_A(t)$ 和上下文 Φ_t。此外，我们还考虑上下文 Φ_t 的变化，其特征由转移概率表示：

$$\Pr\left(\Phi_t = v_r \,|\Phi_{t+1} = v_k\right) = p_{k,r}. \tag{2}$$

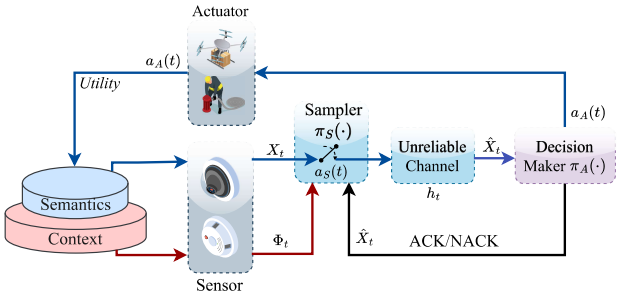


图1. 所考虑系统的示意图，其中传输的语义状态到达接收端，以进行决策，实现某一特定目标。

一般而言，语义与上下文的动力学彼此独立。

B. 不可靠信道与估计转移

我们假设信道实现随时隙变化，彼此独立且同分布（i.i.d.），并遵循伯努利分布。具体而言，在成功传输时，信道实现 h_t 取值为 1，否则取值为 0。因此，我们将成功传输的概率定义为 $\Pr(h_t = 1) = p_S$，将失败概率定义为 $\Pr(h_t = 0) = 1 - p_S$。为刻画 X_t 的动态过程，我们考虑以下两种情况：

- $a_S(t) = 0$。在这种情况下，采样器和发射端均保持空闲状态，这意味着没有新的信息提供给接收端，即 $X_{t+1} = X_t$。因此，我们有：

$$\Pr\left(\hat{X}_{t+1} = x \,\middle|\, \hat{X}_t = s_j, a_S(t) = 0\right) = \mathbb{1}_{\{x = s_j\}}. \tag{3}$$

- $a_S(t) = 1$。在这种情况下，采样器和发射端通过不可靠信道传输当前语义状态 X_t 。由于信道不可靠，我们区分两种不同的情况： $h_t = 1$ 和 $h_t = 0$:

(a) $h_t = 1$。在这种情况下，传输成功。因此，接收端的估计 X_{t+1} 仅为 $X(t)$，转移概率为

$$\Pr\left(\hat{X}_{t+1} = x \,\middle|\, \hat{X}_t = s_j, X_t = s_i, a_S(t) = 1, h_t = 1\right) = \mathbb{1}_{\{x = s_i\}}. \tag{4}$$

(b) $h_t = 0$。在这种情况下，接收端未能成功解码传输内容。因此，接收端的估计值 $\hat{X}_{t+1}$ 保持为 $X(t)$。由此，转移概率为

$$\Pr\left(\hat{X}_{t+1} = x \,\middle|\, \hat{X}_t = s_j, X_t = s_i, a_S(t) = 1, h_t = 0\right) = \mathbb{1}_{\{x = s_j\}}. \tag{5}$$

由于信道实现 h_t 与过程 $\hat{X}_t$、 X_t 和 $a_S(t)$ 相互独立，因此我们有

$$\begin{aligned}
&\Pr\left(\hat{X}_{t+1} = x \,\middle|\, \hat{X}_t = s_j, X_t = s_i, a_S(t) = 1\right) \\
&= \sum_{h_t} p\left(h_t\right) \Pr\left(\hat{X}_{t+1} = x \,\middle|\, \hat{X}_t = s_j, X_t = s_i, a_S(t) = 1, h_t\right) \\
&= p_S \cdot \mathbb{1}_{\{x = s_i\}} + (1 - p_S) \cdot \mathbb{1}_{\{x = s_j\}}.
\end{aligned} \tag{6}$$

结合式 (3) 和式 (6)，可得估计的动力学。

C. 目标导向的决策与执行

我们注意到，先前的研究主要关注如何最小化发射机-接收机系统中的开环新鲜度相关惩罚或误差相关惩罚。然而，无论是新鲜交付，还是准确及时的端到端重建，这些优化工作的最终目标都是确保决策精准有效。为此，我们将开环发射机-接收机信息流拓展为包含决策和执行过程的感知-执行闭环效用流。因此，决策和执行能够将状态更新转化为最终效能。此处，时隙 t 的决策遵循 $a_A(t) = \pi_A(X_t)$，其中 π_A 表示确定性决策策略。

D. 度量：通过 GoT 表征目标

三维 GoT 可由映射 [1]定义：

$$(X_t, \Phi_t, \hat{X}_t) \in \mathcal{S} \times \mathcal{V} \times \mathcal{S} \xrightarrow{\mathcal{L}} \mathrm{GoT}(t) \in \mathbb{R}. \tag{7}$$

在这方面，GoT 记作 $\mathcal{L}(X_t, \Phi_t, \hat{X}_t)$ 或 $\mathrm{GoT}(t)$，表示系统在时隙 t 的瞬时成本，其中 $\hat{}$ 表示关于 (X_t, Φ_t, X_t) 的信息。由 [14],，我们已证明$\hat{}$，给定特定的三元组 (X_t, X_t, Φ_t) 和决策策略 π_A，GoT 可计算如下：

$$\begin{aligned} \mathrm{GoT}^{\pi_A}(t) &= \mathcal{L}(X_t, \Phi_t, \hat{X}_t, \pi_A) \\ &= \Big[C_1(X_t, \Phi_t) - C_2(\pi_A(\hat{X}_t))\Big]^+ + C_3(\pi_A(\hat{X}_t)), \end{aligned} \tag{8}$$

其中，状态固有成本 $C_1(X_t, \Phi_t)$ 用于量化不存在外部影响时，不同语义-上下文对 (X_t, Φ_t) 下的固有成本；执行增益成本 $\hat{C}_2(\pi_A(X_t))$ 用于量化执行 $\pi_A(X(t))$ 所带来的严重程度预期降低 $\hat{}$；执行固有成本 $\hat{}$ $C_3(\pi_A(X_t))$ 则反映特定执行 $\pi_A(\hat{X}(t))$ 所消耗的资源。斜坡函数 $[\cdot]^+$ 确保任何执行 $\pi_A(\hat{X}_t)$ 都能将成本降低至最高为 0。图 2 展示了 GoT 构造的一个示例。图 2 中的 GoT 根据以下定义得到：

$$\begin{aligned} &C_1(X_t, \Phi_t) = \left(\begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 0 & 1 & 3 \\ 1 & 0 & 2 & 5 \end{array}\right), \quad \pi_A(\hat{X}_t) = [0, 1, 2], \\ &C_2(\pi_A(\hat{X}_t)) = 2\pi_A(\hat{X}_t), \qquad\qquad C_3(\pi_A(\hat{X}_t)) = \pi_A(\hat{X}_t). \end{aligned} \tag{9}$$

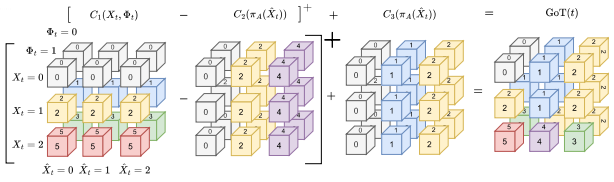
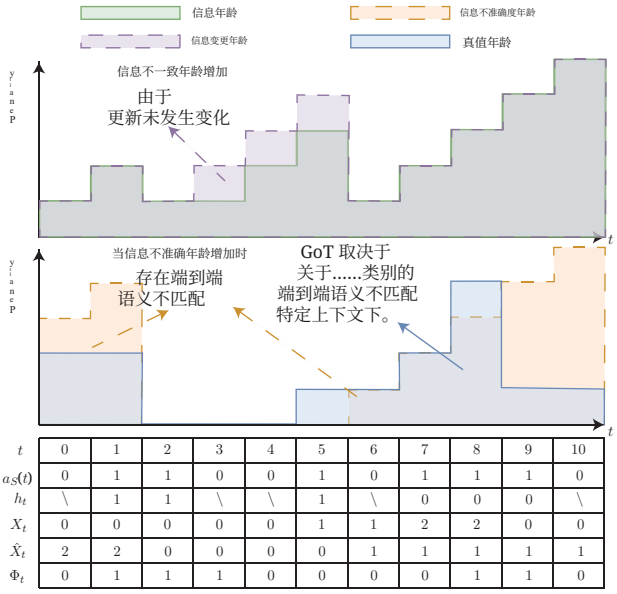


图 2. 通过式 (8) 和 (9) 表征 GoT 的可视化示例。

[1]值得注意的是，通过整合执行策略、任务特定属性及其他相关因素，GoT 可以扩展到更高维度。



t	0	1	2	3	4	5	6	7	8	9	10
$a_S(t)$	0	1	1	0	0	1	0	1	1	1	0
h_t	\	1	1	\	\	1	\	0	0	0	\
X_t	0	0	0	0	0	1	1	2	2	0	0
$\hat{X}_t$	2	2	0	0	0	0	1	1	1	1	1
Φ_t	0	1	1	1	0	0	0	0	1	1	0

图 3. 时隙状态更新系统中的信息年龄、信息不一致年龄、信息不准确年龄和 GoT 示意图。这里，GoT 的值取自图 2 右侧的张量。

图 3 展示了信息年龄、信息不一致年龄、信息不准确年龄和 GoT 的瞬时变化过程。从图 3 中的时隙 $t = 0$, 1，9，10 可以明显看出信息不准确年龄固有的局限性：不匹配持续一段时间，并不一定会导致成本增加。相反，端到端语义不匹配的类别有助于体现真实的瞬时成本。

III. 问题建模与求解

传统上，采样策略的制定与决策过程彼此独立。两阶段方法的典型示例是：首先根据信息年龄或其变体确定最优采样策略，然后再进行有效决策。这种两阶段分离设计源于现有指标的固有限制，即它们无法捕捉闭环决策效用。然而，GoT 指标使我们能够协同设计采样与决策。我们探讨团队决策理论，其中两个智能体——一个作为采样器，另一个作为决策者——协作以实现共同目标。我们的目标是确定一个联合确定性策略 $\boldsymbol{\pi}_C = (\pi_S, \pi_A)$，使系统的长期平均成本最小化。我们认为，采样和传输更新也会消耗能量，从而产生 C_s 成本。在这种情况下，系统的瞬时成本可由 $\mathrm{GoT}^{\pi_A}(t) + C_s \cdot a_S(t)$明确表示，问题可表述为：

$$\mathcal{P}1: \quad \min_{\boldsymbol{\pi}_C \in \Upsilon} \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}^{\boldsymbol{\pi}_C} \left(\sum_{t=0}^{T-1} \mathrm{GoT}^{\pi_A}(t) + C_s \cdot a_S(t) \right), \tag{10}$$

其中，$\boldsymbol{\pi}_C = (\pi_S, \pi_A)$ 表示联合采样与决策策略，由 $\pi_S = (a_S(0)\ a_S(1)\ \cdots)$ 和$\pi_A = (a_A(0)\ a_A(1)\ \cdots)$ 组成，分别对应采样

动作序列和执行序列。注意，GoT$^{\pi_A}(t)$ 由式 (8) 表征。

A. Dec-POMDP 表述

为了解决问题 $\mathcal{P}1$，我们应将其表述为一个分散式部分可观测马尔可夫决策过程（DEC-POMDP）问题，该问题最初在 [15]中提出，用于解决分布式多智能体的协作序贯决策问题。在 Dec-POMDP 框架中，一组智能体仅依靠各自的局部知识进行协作，以实现共同目标。典型的 Dec-POMDP 可表示为元组 $\mathscr{M}_{DEC-POMDP} \triangleq \langle n, \mathcal{I}, \mathcal{A}, \mathcal{T}, \Omega\ \mathcal{O},\ \mathcal{R}\rangle$:

- n 表示智能体的数量。在所考虑的模型中，我们有 $n = 2$，这意味着存在两个智能体：一个智能体 $\mathcal{A}gent_S$ 作为采样器，另一个则代表决策者，记为 $\mathcal{A}gent_A$。

- $\mathcal{I}$ 是全局系统状态的有限集合，其特征由 $(X_t,\ X_t,\ \Phi_t) \in \mathcal{S} \times \mathcal{S} \times \mathcal{V}$ 给出。为简洁起见，以下将$\mathbf{W}_t = (X_t,\ \hat{}X_t,\ \Phi_t)$记作简写。

- $\mathcal{T}$是由以下公式定义的转移函数

$$\mathcal{T}(\mathbf{w}, \mathbf{a}, \mathbf{w}') \triangleq \Pr(\mathbf{W}_{t+1} = \mathbf{w}' | \mathbf{W}_t = \mathbf{w}, \mathbf{a}_t = \mathbf{a}), \quad (11)$$

其定义为系统中的智能体执行联合动作$\mathbf{a}_t = \mathbf{a} = (a_S(t)\ a_A(t))$后，从全局状态$\mathbf{W}_t = \mathbf{w}$ 转移至状态$\mathbf{W}_{t+1} = \mathbf{w}'$的转移概率。为简化记号，在下文中我们令 $p(\mathbf{w}'|\mathbf{w},\mathbf{a})$表示$\mathcal{T}(\mathbf{w},\mathbf{a},\mathbf{w}')$。然后，根据引理1可计算转移函数

引理1。 分散式部分可观测马尔可夫决策过程的转移函数:

$$\begin{aligned} &p\left((s_u, x, v_r)\,|(s_i, s_j, v_k), (1, a_m)\right) = \\ &p_{i,u}^{(k,m)} \cdot p_{k,r} \cdot \left(p_S \cdot \mathbb{1}_{\{x=s_i\}} + (1 - p_S) \cdot \mathbb{1}_{\{x=s_j\}}\right) \end{aligned}, \quad (12)$$

$$\begin{aligned} &p\left((s_u, x, v_r)\,|(s_i, s_j, v_k), (0, a_m)\right) = \\ &p_{i,u}^{(k,m)} \cdot p_{k,r} \cdot \mathbb{1}_{\{x=s_j\}} \end{aligned}, \quad (13)$$

对于任意 $x \in \mathcal{S}$以及索引 i、j、$u \in \{1, 2, \cdots, |\mathcal{S}|\}$、$k$、$r \in \{1, 2, \cdots, |\mathcal{V}|\}$和 $m \in \{1, 2, \cdots, |\mathcal{A}_A|\}$。

证明。 考虑到在给定（X_t, Φ_t, X_t）和$\mathbf{a}(t)$的情况下，X_{t+1}、Φ_{t+1}与 X_{t+1}之间具有条件独立性，将公式（1）、（2）、（3）和（6）中的动力学纳入其中，即可推导出转移函数。□

- $\mathcal{A} = \mathcal{A}_S \times \mathcal{A}_A$，其中 $\mathcal{A}_S \triangleq \{0, 1\}$ 表示采样器的动作集，$\mathcal{A}_A \triangleq \{a_0, \cdots, a_{M-1}\}$ 表示决策者的动作集。

- $\Omega = \Omega_S \times \Omega_A$，其中 Ω_S 表示采样器的观测域。在此情况下，采样器 $\mathcal{A}gent_S$ 完全可观测，Ω_S 涵盖完整的系统状态 $o_{(t)\ S} = \mathbf{W}_t$。$\Omega_A$ 表示执行器的观测域。在此情况下，决策者 $\mathcal{A}gent_A$ 是部分可观测的，Ω_A 由 $o_{(t)\ A} = \hat{}X(t)$ 组成。时刻 t 的联合观测表示为 $\mathbf{o}_t = (o_{(t)\ S}, o_{(t)\ A})$。

- $\mathcal{O} = \mathcal{O}_S \times \mathcal{O}_A$ 表示观测函数，其中$\mathcal{O}_S$ 和 $\mathcal{O}_A$ 分别表示采样器$\mathcal{A}gent_S$ 和执行器 $\mathcal{A}gent_A$的观测函数，定义如下:

$$\begin{aligned} \mathcal{O}(\mathbf{w}, \mathbf{o}) &\triangleq \Pr(\mathbf{o}_t = \mathbf{o} | \mathbf{W}_t = \mathbf{w}), \\ \mathcal{O}_S(\mathbf{w}, o_S) &\triangleq \Pr(o_S^{(t)} = o_S | \mathbf{W}_t = \mathbf{w}), \\ \mathcal{O}_A(\mathbf{w}, o_A) &\triangleq \Pr(o_A^{(t)} = o_A | \mathbf{W}_t = \mathbf{w}). \end{aligned} \quad (14)$$

智能体 $\mathcal{A}gent_i$ 的观测函数表示在当前全局系统状态为 $\mathbf{W}_t$ =**w**的条件下，智能体 $\mathcal{A}gent_i$ 感知到 o_i的条件概率。为简洁起见，下文将 $p_A(o_A|\mathbf{w})$ 表示为$\mathcal{O}_A(\mathbf{w}\ o_A)$，并将 $p_S(o_S|\mathbf{w})$ 表示为 $\mathcal{O}_S(\mathbf{w}\ o_A)$。在我们所考虑的模型中，观测函数是确定性的，其特征由引理 2给出。

引理 2. 分散式部分可观测马尔可夫决策过程的观测函数:

$$\begin{aligned} p_S\left((s_u, s_r, v_q)|(s_i, s_j, v_k)\right) &= \mathbb{1}_{\{(s_u,s_r,v_q)=(s_i,s_j,v_k)\}}, \\ p_A\left(s_z|(s_i, s_j, v_k)\right) &= \mathbb{1}_{\{s_z=s_j\}}. \end{aligned} \quad (15)$$

对于索引 z、i、j、u、$r \in \{1, 2, \cdots |\mathcal{S}|\}$ 以及 k、$q \in \{1, 2, \cdots |\mathcal{V}|\}$。

- $\mathcal{R}$ 是奖励函数，其特征是映射$\mathcal{I} \times \mathcal{A} \to \mathbb{R}$。在最大化长期平均奖励的设定下，解决分散式部分可观测马尔可夫决策过程等价于处理以下问题:

$$\min_{\boldsymbol{\pi}_C \in \Upsilon} \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}^{\boldsymbol{\pi}_C} \left(-\sum_{t=0}^{T-1} r(t) \right). \quad (16)$$

随后，为了与问题$\mathcal{P}1$保持一致，相应地将奖励函数定义为:

$$r(t) = \mathcal{R}^{\pi_A}(\mathbf{w}, a_S) = -\mathrm{GoT}^{\pi_A}(t) - C_s \cdot a_S(t). \quad (17)$$

B. 无限时域分散式部分可观测马尔可夫决策过程的求解

一般而言，已知在有限时域设定 [15]下，求解分散式部分可观测马尔可夫决策过程是 NEXP 完全问题。对于无限时域分散式部分可观测马尔可夫决策过程问题，已知寻找其最优策略是不可判定的。然而，在我们考虑的模型中，采样和决策过程均被视为确定性的，分别表示为 $a_S(t) = \pi_S(\mathbf{w})$ 和 $a_A(t) = \pi_A(o_A)$。在这种情况下，可以通过在决策策略空间中进行暴力搜索，确定一个联合最优策略。

其核心思想基于这样一个发现：给定一个确定性决策策略 π_A，采样问题可以被表述为一个标准的全观测马尔可夫决策过程（MDP）问题，记为 $\mathscr{M}^{\pi_A}{}_{\mathrm{MDP}} \triangleq \langle \mathcal{I}, \mathcal{T}^{\pi_A}, \mathcal{A}_S, \mathcal{R}\rangle$。

命题 1。 给定一个确定性决策策略 π_A，最优采样问题可以表述为一个典型的全观测马尔可夫决策过程问题 $\mathscr{M}^{\pi_A}{}_{\text{马尔可夫决策过程}} \triangleq \langle \mathcal{I}, \mathcal{A}_S, \mathcal{T}^{\pi_A}{}_{\text{马尔可夫决策过程}}\ \mathcal{R}\rangle$，其要素如下:

- $\mathcal{I}$：与预定义的*Dec-POMDP*元组相同。
- $\mathcal{A}_S = \{0, 1\}$：采样与传输动作集。

• $\mathcal{T}^{\pi_A}$：给定确定性决策策略 π_A 时的转移函数，其为

$$\begin{aligned}\mathcal{T}^{\pi_A}(\mathbf{w}, a_S, \mathbf{w}') &= p^{\pi_A}(\mathbf{w}'|\mathbf{w}, a_S) \\ &= \sum_{o_A \in \mathcal{O}_A} p\left(\mathbf{w}'|\mathbf{w}, (a_S, \pi_A(o_A))\right) p_A(o_A|\mathbf{w}),\end{aligned} \tag{18}$$

其中，$p\left(\mathbf{w}'|\mathbf{w},(a_S,\pi_A(o_A))\right)$ 可由引理*1*得出，而 $p(o_A|\mathbf{w})$可由引理*2*得出。

• $\mathcal{R}$：与预定义的*Dec-POMDP*元组相同。

现在我们着手求解MDP问题 $\mathscr{M}^{\pi_A}_{\text{MDP}}$。为了推导决策策略 π_A下的最优采样策略，必须求解贝尔曼方程 [16]:

$$\begin{aligned}&\theta^*_{\pi_A} + V_{\pi_A}(\mathbf{w}) = \\ &\max_{a_S \in \mathcal{A}_A} \left\{ \mathcal{R}^{\pi_A}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}'|\mathbf{w}, a_S) V_{\pi_A}(\mathbf{w}') \right\},\end{aligned} \tag{19}$$

其中，V^{π_A}(**w**)为价值函数，$\theta^*_{\pi_A}$ 是决策策略 π_A下的最优长期平均奖励。我们采用相对价值迭代（RVI）算法求解该问题。具体步骤如算法1所示：

Algorithm 1: The RVI Algorithm to Solve the MDP Given the decision policy π_A

Input: The MDP tuple $\mathscr{M}^{\pi_A}_{\text{MDP}}$, ϵ, π_A;

Initialization: $\forall \mathbf{w} \in \mathcal{I}$, $\tilde{V}^0_{\pi_A}(\mathbf{w}) = 0$, $\tilde{V}^{-1}_{\pi_A}(\mathbf{w}) = \infty$, $k = 0$;

Choose $\mathbf{w}^{ref}$ arbitrarily;

while $||\tilde{V}^k_{\pi_A}(\mathbf{w}) - \tilde{V}^{k-1}_{\pi_A}(\mathbf{w})|| \geq \epsilon$ **do**

  $k = k + 1$;

  for $\mathbf{w} \in \mathcal{I} - \mathbf{w}^{ref}$ **do**

    $\tilde{V}^k_{\pi_A}(\mathbf{w}) = -g_k +$

    $\max_{a_S} \left\{ \mathcal{R}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I} - \mathbf{w}^{ref}} p(\mathbf{w}'|\mathbf{w}, a_S) \tilde{V}^{k-1}_{\pi_A}(\mathbf{w}') \right\}$;

$\theta^*(\pi_A, \pi^*_S) = -\tilde{V}^k_{\pi_A}(\mathbf{w})$

$\max_{a_S \in \mathcal{A}_S} \left\{ \mathcal{R}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}'|\mathbf{w}, a_S) \tilde{V}^k_{\pi_A}(\mathbf{w}') \right\}$;

for $\mathbf{w} \in \mathcal{I}$ **do**

  $\pi^*_S(\pi_A, \mathbf{w}) = \arg\max_{a_S} \left\{ \mathcal{R}(\mathbf{w}, a_S) + \sum_{\mathbf{w}' \in \mathcal{I}} p(\mathbf{w}'|\mathbf{w}, a_S) \tilde{V}^k_{\pi_A}(\mathbf{w}') \right\}$;

Output: $\pi^*_S(\pi_A)$, $\theta^*(\pi_A, \pi^*_S)$

借助命题 1和算法1，我们可以在决策策略空间 Υ_A 中进行暴力搜索，从而获得联合采样与决策策略。该算法称为RVI 暴力搜索算法，详见算法2。

定理1。在转移函数 $\mathcal{T}^{\pi_A}$ 遵循单链的情况下，*RVI* 暴力搜索算法（算法*2*）能够实现最优联合确定性策略（π^*_S、π^*_A）。

算法2： RVI 暴力搜索算法

输入： 分散式部分可观测马尔可夫决策过程元组 $\mathscr{M}_{DEC-POMDP}$;

对 $\pi_A \in \Upsilon$ **执行**

  构建MDP问题 $\mathscr{M}_{\text{MDP}} = \langle \mathcal{I}, \mathcal{A}_S, \mathcal{T}^{\pi_A}{}_{\text{MDP}}, \mathcal{R} \rangle$ 如命题 1;

  运行算法1以获得 $\pi^*_S(\pi_A)$和 $\theta^*(\pi_A, \pi^*_S)$;

计算最优联合策略:

$$\begin{cases} \pi^*_A = \arg\min_{\pi_A} \theta^*_{\pi_A}; \\ \pi^*_S = \pi_S(\pi^*_A) \end{cases}$$

Output: π^*_S, π^*_A

证明。如果转移函数 $\mathcal{T}^{\pi_A}$ 遵循单链，则根据 [17, 定理 8.4.5] 可知，对于任意 π_A，我们可以获得最优确定性策略 π^*_S，使得$\theta^*(\pi_A, \pi^*_S) \leq \theta^*(\pi_A, \pi_S)$。此外，算法2保证，对于任意 π_A，$\theta^*(\pi^*_A, \pi^*_S) \leq \theta^*(\pi_A, \pi^*_S)$。由此可以得出结论，对于任意 $\boldsymbol{\pi}_C = (\pi_S, \pi_A) \in \Upsilon$，我们有

$$\theta^*(\pi^*_A, \pi^*_S) \leq \theta^*(\pi_A, \pi_S). \tag{20}$$

IV. 仿真结果

对于仿真设置，我们设定 $\mathcal{A}_A = \{0, \cdots, 10\}$、$\mathcal{S} = \{s_0, s_1, s_2\}$、$\mathcal{V} = \{v_0, v_1, v_2\}$，相应的成本为:

$$\mathbf{C}_1(X_t, \Phi_t) = \begin{pmatrix} 0 & 20 & 50 \\ 0 & 10 & 20 \end{pmatrix}, \tag{21}$$

我们假设 $C_2(\pi_A(\hat{}X_t))$ 和 $C_3(\pi_A(\hat{}X_t))$ 均关于 $\hat{}$ 决策呈线性关系，其中 $C_2(\pi_A(X_t) = C_g \cdot \pi_A(X_t)$ 和 $C_3(\pi_A(\hat{}X_t)) = C_I \cdot \pi_A(\hat{}X_t)$，且 $C_g = 8$ 和 $C_I = 1$。

A. 基准比较：分离设计

对于决策，我们考虑决策策略 π_A 由贪心方法预先确定:

$$\begin{aligned}\pi_A(\hat{X}_t) = \arg\min_{a_A \in \mathcal{S}_A} \mathop{\mathbb{E}}_{\Phi_t} \Big\{ &\Big[C_1(\hat{X}_t, \Phi_t) - C_2(\pi_A(\hat{X}_t)) \Big]^+ \\ &+ C_3(\pi_A(\hat{X}_t)) \Big\}.\end{aligned} \tag{22}$$

这种基于贪心的方法是在假设 $\hat{}$估计 X_t 准确的情况下，选择使当前步骤成本最小的决策。通过计算（22），我们得到基于贪心的决策策略 $\pi_A(\hat{}X_t) = [0, 3, 7]$。在此决策策略下，采样设计考虑以下采样基准:

• **均匀采样**。采样按周期触发，即，$a_S(t) = \mathbb{1}_{\{t = K * \Delta\}}$，其中 $K = 0, 1, 2, \cdots$ 和 $\Delta \in \mathbb{N}^+$。对于每个 Δ，计算采样率为 $1/\Delta$，并通过马尔可夫链仿真获得长期平均成本。

• **年龄感知**。当信息年龄达到预设阈值时执行采样，即，$a_S(t) = \mathbb{1}_{\{\text{信息年龄} > \delta\}}$，其中信息年龄最优阈值 δ 可通过 [5]中算法1所述的二分法确定。在图

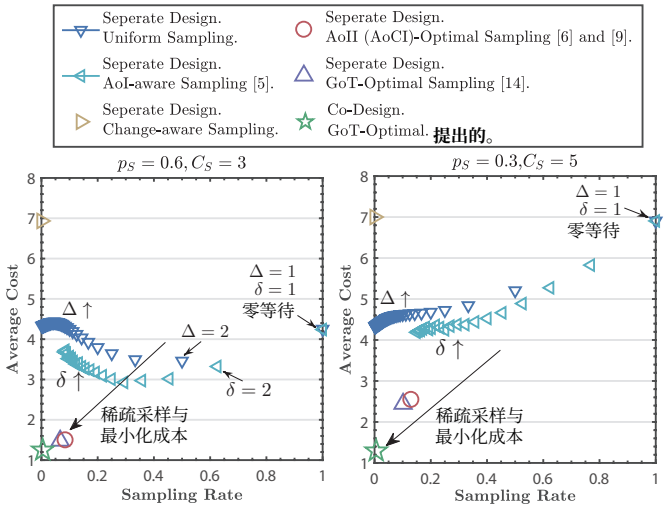


图 4. 不同策略和参数设置下的平均成本与采样率关系。均匀采样和信息年龄感知策略的曲线通过调整间隔 Δ 和 δ 得到。

4中，我们动态调整阈值 δ，以探索采样率与效用之间的平衡。

- **变化感知**采样会在源状态发生变化时触发，即，$a_S(t) = \mathbb{1}_{\{X_t \neq X_{t-1}\}}$。该策略的采样频率深受系统动力学的影响：如果源频繁传输，采样率就会较高；而如果传输次数较少，采样率就会较低。

- **最优 AoII（也包括最优 AoCI）**。根据 [9], , 已证明 AoII 最优采样策略为$a_S(t) = \mathbb{1}_{\{X_t \neq \hat{X}_t\}}$。根据 [6], , AoCI 最优采样策略为 $a_S(t) = \mathbb{1}_{\{X_t \neq X_{t-\mathrm{AoI}(t)}\}}$。请注意，$X_t = X_{t-\mathrm{AoI}(t)}$，这两种采样策略是等价的。给定此采样策略和基于贪心的决策策略，可以得到采样率和平均成本。

B. 通过 *GoT* 实现协同设计

我们注意到，采样与决策密切交织，这凸显了进一步探索联合设计的潜力。本文提出了 RVI-暴力搜索算法（算法2），以分布式方式求得最优联合策略。如图4所示，采样器—决策者 协同设计仅通过稀疏采样便可实现最优的长期平均效用。只有对决策至关重要、承载关键信息语义的内容才会被采样和传输，其他内容则被过滤。通过采用最佳匹配决策策略，所提出的目标导向、语义感知和稀疏采样方法比现有方法表现更优。因此，目标导向、语义与上下文感知的稀疏采样旨在通过有效决策实现最大效用。

V. 结论

本文研究了 GoT 度量，以直接描述目标导向系统的决策制定效用效用。借助 GoT，我们构建了一个无限时域分散式部分可观测马尔可夫决策过程问题，以实现采样与决策的集成设计。为解决这一问题，我们开发了 RVI 暴力搜索算法，以求得最优解。对比分析证实，所提出的 GoT 最优采样器-决策者协同设计能够实现稀疏采样，同时最大化效用，从而实现稀疏采样器与目标导向决策制定者的协同设计。

参考文献

[1] S. Kaul、M. Gruteser、V. Rai 和 J. Kenney，“车载网络中信息年龄的最小化”，载于*2011*年第*8*届*IEEE*通信学会传感器、网状及自组织通信与网络年会。IEEE，2011年，第350–358页。[2] R. D. Yates、Y. Sun、D. R. Brown、S. K. Kaul、E. Modiano 和 S. Ulukus，“信息年龄：介绍与综述”，*IEEE*通信领域精选期刊，第39卷，第5期，第1183–1210页，2021年。[3] A. Kosta、N. Pappas、A. Ephremides 和 V. Angelakis，“信息年龄与信息价值：非线性年龄情形”，载于*2017*年*IEEE*信息论国际研讨会（*ISIT*）。IEEE，2017年，第326–330页。[4] ——，“状态更新的延迟成本及其价值：非线性老化”，*IEEE*通信学报，第68卷，第8期，第4905–4918页，2020年。[5] Y. Sun 和 B. Cyr，“数据新鲜度优化的采样：非线性年龄函数”，通信与网络期刊，第21卷，第3期，第204–219页，2019年。[6] X. Wang、W. Lin、C. Xu、X. Sun 和 X. Chen，“变更信息年龄：物联网中的内容感知状态更新”，*IEEE*通信学报，第70卷，591，2022年1，578期，第–页。[7] J. Zhong、R. D. Yates 和 E. Soljanin，“本地缓存刷新的两种新鲜度指标”，载于*2018*年*IEEE*信息论国际研讨会（*ISIT*）。IEEE，2018年，第1924–1928页。[8] X. Zheng、S. Zhou 和 Z. Niu，“远程控制系统中上下文感知及时状态更新的信息紧迫性”，*IEEE*无线通信学报，第19卷，7250，2020年11，7237期，第–页。[9] A. Maatouk、S. Kriouile、M. Assaad 和 A. Ephremides，“错误信息年龄：一种新的状态更新性能指标”，*IEEE/ACM*网络学报，第28卷，2228，2020年5，2215期，第–页。[10] M. Kountouris 和 N. Pappas，“面向联网智能系统的语义赋能通信”，*IEEE*通信杂志，第59卷，第96–102页，2020年。[11] N. Pappas 和 M. Kountouris，“自主系统中面向实时跟踪的目标导向通信”，载于*2021*年*IEEE*自主系统国际会议（*ICAS*），2021年，第1–5页。[12] M. Salimnejad、M. Kountouris 和 N. Pappas，“无线信道上的马尔可夫源实时重构与远程执行”，*arXiv*，2023年。[在线]。网址：https://arxiv.org/abs/ 2302 13927 。E. [13] Fountoulakis、N. Pappas 和 M. Kountouris，“无线自主系统中执行误差成本最小化的目标导向策略”，*arXiv*，2023年。 [在线]。网址：https://arxiv.org/abs/2303. 04908[14]A. Li、S. Wu、S. Meng 和 Q. Zhang，“迈向目标导向语义通信：新指标、开放性挑战与未来研究方向”，已提交至*IEEE*通信杂志。，2023 [ ] 2304 00848 。在线。网址：[15] https://arxiv.org/abs/。D. S. Bernstein、R. Givan、N. Immerman 和 S. Zilberstein，“马尔可夫决策过程分散式控制的复杂性”，运筹学数学，第27卷，第4期，第819–840页，2002年。[16] D. Bertsekas，动态规划与最优控制：第一卷。Athena Scientific，2012年，第1卷。[17] M. L. Puterman，马尔可夫决策过程：离散随机动态规划。John Wiley & Sons，2014年。