arXiv:2301.00621v1 [cs.IT] 2 Jan 2023

离散字母表上定向信息的数据驱动优化

Dor Tsur*、Ziv Aharoni*、Ziv Goldfeld† 和 Haim Permuter*

摘要

定向信息（DI）是研究和分析序贯随机模型的一项基本度量。具体而言，对输入分布进行优化时，它可表征一般通信信道的容量。然而，DI 的解析计算通常难以处理，而现有针对离散输入字母表的优化技术需要了解信道模型，因此在仅有样本可用时无法应用。为克服这些局限，我们提出一种针对离散输入空间中 DI 的新型估计优化框架。我们将 DI 优化建模为马尔可夫决策过程，并运用强化学习技术优化输入过程概率质量函数（PMF）的深度生成模型。将该优化器与近期开发的 DI 神经估计器相结合，我们得到一种端到端估计优化算法，并将其用于估计各种有记忆离散信道的（前向和反馈）容量。此外，我们展示了如何使用优化后的 PMF 模型来：（i）获得单值有限状态信道反馈容量的理论界；以及（ii）在峰值功率受限加性白高斯噪声信道中实现星座的概率整形。

1 引言

定向信息（DI）最初是为研究带反馈信道而提出的，用于量化随机过程之间的统计和时间依赖关系。 [1]它已广泛应用于通信领域[2], 投资组合理论 [3], 计算生物学 [4],神经科学 [5] 和机器学习 [6, 7]。通常，人们希望针对某些相关随机过程的分布优化互信息；例如，带反馈和不带反馈时的信道容量，分别由在适当输入分布集合上最大化的互信息给出 [8]。然而，由此产生的优化问题通常难以处理，只有有限类别的信道存在解析解 [9–12]。

在缺乏解析解的情况下，可以通过互信息优化方法计算容量，例如 Blahut-Arimoto型算法 [13, 14] ，或基于马尔可夫决策过程（MDP）表述和动态规划 [10, 15] ，或强化学习 [16] 技术的方法。然而，这些方法只有在信道满足严格的结构假设时才可行，

*内盖夫本-古里安大学
†康奈尔大学

这些假设使多字母互信息目标能够张量化，例如单叶[1] 具有反馈的有限状态信道（FSC）和对称信道。对于更一般的信道[2], , 可以利用 Q-图 [18, 19], 工具来界定容量，但要获得紧界，就需要在指数规模的空间中进行穷举搜索。此外，所有上述方法都需要完整了解信道概率模型，而这在实践中往往无法获得。

当信道模型未知但可以从中采样时，可以采用经验DI估计器来获得近似解。对于无记忆信道， [20, 21]提出了在连续输入空间上联合估计与优化的方法。近期，[22] 利用其中提出的DI神经估计器（DINE）处理了有记忆信道的情形。DINE使用循环神经网络（RNN）对互信息的Donsker-Varadhan表示进行参数化，以样本均值近似期望，并在参数空间中优化由此得到的目标函数。为了计算反馈容量， [22] 进一步提出了一种用于连续输入分布的基于RNN的生成模型，并通过在两个模型中传播梯度，对其与DINE进行联合优化。这些方法依赖联合模型端到端的可微性，而离散输入字母表不具备这一性质。据我们所知，迄今尚无已知的数据驱动方法能够针对离散字母表优化（估计的）互信息。本文旨在弥补这一空白。

我们提出了一种在离散输入字母表上优化 DINE 的新方法。输入分布由基于 RNN 的概率质量函数（PMF）生成器建模。我们将 DI最大化表述为一个马尔可夫决策过程，其策略由概率质量函数生成器建模。这种表述使我们能够利用强化学习技术，尤其是通过策略梯度定理进行策略优化 [23]。结合基于 DINE 的马尔可夫决策过程奖励近似以及策略梯度表达式的蒙特卡罗（MC）估计，这一结果形成了一个可处理的策略优化目标。随后，我们交替优化这一目标和 DINE参数，从而得到一种针对离散输入信道 DI速率的估计-优化过程。重要的是，我们的方法不依赖于信道转移核的任何先验知识，而只要求能够对其输出进行采样。

我们将所提出的方法应用于关于噪声信道通信的三项主要任务。首先，我们使用该方法估计多个有记忆信道的容量。在所有所考察的示例中，该方法要么达到理论容量值，要么收敛于已知上界和下界之间。此外，我们证明，优化的 PMF 生成器与已知的达到容量的输入分布相对应。其次，我们利用该生成器估计一个 Q-图 [24], , 该图可插入 [18] 中的算法，以获得单叶有限状态信道反馈容量的紧界。最后，我们利用所开发的方法，对峰值功率受限的加性白高斯噪声（AWGN）信道上的脉冲幅度调制（PAM）和正交幅度调制（QAM）星座执行概率整形。我们的方法得出

[1]如果有限状态信道的状态演化是过去的输入、输出和状态元组的时不变确定性函数，则称其为单叶。[2]或者当反馈容量本身无法通过算法（博雷尔-图灵）计算时 [17]。

非平凡分布，其信息传输率高于均匀分布所获得的传输率，而均匀分布通常在实践中使用 [25]。

本文其余部分安排如下。第2 节介绍预备知识和技术背景。第3 节推导定向信息优化器，第4节讨论实现和算法方面的问题。第5 节展示信道容量估计的实证结果。第6 节和第7节分别重点讨论单叶有限状态信道反馈容量的界定技术及其在概率整形中的应用。证明见第8节。第9 节总结全文，并讨论未来研究方向。

2 背景与预备知识

2.1 符号说明

集合用花体字母表示，例如 $\mathcal{X}$。当 $\mathcal{X}$ 是有限集时，我们用 $|\mathcal{X}|$ 表示其基数。对于任意 $n \in \mathbb{N}$，$\mathcal{X}^n$ 是 $\mathcal{X}$ 的 n 重笛卡尔积，而 $x^n = (x_1, \ldots, x_n)$ 表示 $\mathcal{X}^n$ 中的一个元素。对于 i、$j \in \mathbb{Z}$，若满足 $i \le j$，我们使用简写 $x_i^j := (x_i, \ldots, x_j)$；当 $i = 1$ 时省略下标。我们用 $(\Omega, \mathcal{F}, \mathbb{P})$ 表示定义所有随机变量的底层概率空间，并假设该空间足够丰富。期望用 $\mathbb{E}$ 表示；有时我们写作 $\mathbb{E}_P$，以强调底层分布为 P。$\mathcal{X}$ 上所有博雷尔概率测度的集合记为 $\mathcal{P}(\mathcal{X})$，k 维概率单纯形记为 Δ_k。当 $\mathcal{X}$ 可数时，我们用 p 表示与 $P \in \mathcal{P}(\mathcal{X})$ 相关的概率质量函数。随机变量用大写字母表示，例如 X；随机过程用黑板粗体字母表示，例如 $\mathbb{X} := (X_i)_{i \in \mathbb{N}}$。

对于满足 $Q \ll P$的 P、$Q \in \mathcal{P}(\mathcal{X})$，即 Q关于 P绝对连续，我们将 P 关于 Q 的拉东-尼科迪姆导数记为 $\frac{dP}{dQ}$。在 $P \ll Q$的条件下，P 与 Q之间的KL散度为$\mathsf{D}_{\mathsf{KL}}(P\|Q) := \mathbb{E}_P\left[\log \frac{dP}{dQ}\right]$。$(X, Y) \sim P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$之间的互信息为$\mathsf{I}(X; Y) := \mathsf{D}_{\mathsf{KL}}(P_{XY}\|P_X \otimes P_Y)$，其中 P_X和 P_Y是 P_{XY}的边缘分布。离散随机变量 $X \sim P$的熵为 $H(X) := -\mathbb{E}[\log p(X)]$。

2.2 定向信息与信道容量

最初由 Massey 提出[1], 定向信息量化了一个随机变量序列以因果方式向另一个序列传递的信息量。

定义 1（定向信息）设 $(X^n, Y^n) \sim P_{X^nY^n} \in \mathcal{P}(\mathcal{X}^n \times \mathcal{Y}^n)$。从 X^n 到 Y^n 的互信息为

$$\mathsf{I}(X^n \to Y^n) := \sum_{i=1}^{n} \mathsf{I}(X^i; Y_i | Y^{i-1}). \tag{1}$$

对于无限时间范围内的联合平稳随机过程 $\mathbb{X}$ 和 $\mathbb{Y}$，二者之间的 DI速率定义为渐近时间平均 DI：

$$\mathsf{I}(\mathbb{X} \to \mathbb{Y}) := \lim_{n \to \infty} \frac{1}{n} \mathsf{I}(X^n \to Y^n).$$

联合平稳性确实足以保证该极限的存在 [26]。

信道序列 $\{P_{Y^n\|X^n}\}_{n\in\mathbb{N}}$ 的前馈容量和反馈容量（其中 $P_{Y^n\|X^n}:=\prod_{i=1}^n P_{Y_i|Y^{i-1}X^i}$）均可表征为[3]

$$C=\lim_{n\to\infty}\sup_{P}\frac{1}{n}\mathsf{I}(X^n\to Y^n), \tag{2}$$

其中 $P=P_{X^n}$ 表示前馈容量，而 $P=P_{X^n\|Y^{n-1}}:=\prod_{i=1}^n P_{X_i|X^{i-1}Y^{i-1}}$（在存在反馈时，称为因果条件分布）。另请注意，当不存在反馈时，我们有 $\mathsf{I}(X^n;Y^n)=\mathsf{I}(X^n\to Y^n)$ [1]。

2.3 互信息神经网络估计

DINE [22] 是一种基于循环神经网络的估计器，可根据样本 $D_n:=(X^n,Y^n)\sim P_{X^nY^n}$ 估计 $\mathsf{I}(\mathbb{X}\to\mathbb{Y})$。其推导首先将 DI 速率表示为以下 KL 散度项之差的渐近形式：

$$\begin{aligned}\mathsf{D}^N_{Y\|X}&:=\mathsf{D}_{\mathsf{KL}}\left(P_{Y^0_{-(N-1)}\|X^0_{-(N-1)}}\middle\|P_{Y^{-1}_{-(N-1)}\|X^{-1}_{-(N-1)}}\otimes P_{\widetilde{Y}}\middle|P_{X^0_{-(N-1)}\|Y^{-1}_{-(N-1)}}\right),\\ \mathsf{D}^N_{Y}&:=\mathsf{D}_{\mathsf{KL}}\left(P_{Y^0_{-(N-1)}}\middle\|P_{Y^{-1}_{-(N-1)}}\otimes P_{\widetilde{Y}}\right),\end{aligned} \tag{3}$$

其中 $\mathsf{D}_{\mathsf{KL}}(P_{Y|X}\|Q_{Y|X}|P_X):=\mathbb{E}_{P_x}[\mathsf{D}_{\mathsf{KL}}(P_{Y|X}\|Q_{Y|X})]$。

该估计器采用 KL散度 的 Donsker-Varadhan (DV) 变分形式[28, 定理 3.2], 据此，对于任意 $P,Q\in\mathcal{P}(\mathcal{X})$，只要满足 $P\ll Q$，就有

$$\mathsf{D}_{\mathsf{KL}}\left(P\|Q\right)=\sup_{f:\mathcal{X}\to\mathbb{R}}\mathbb{E}_P\left[f\right]-\log\left(\mathbb{E}_Q[e^f]\right). \tag{4}$$

上确界是在所有可测函数 f 中取的，这些函数的期望有限（称为 DV势函数），且上确界由 $f^\star=\log\frac{dP}{dQ}+c$，对于任意 $c\in\mathbb{R}$。将 DV 公式（4）应用于（3）中的 KL散度，可得

$$\mathsf{D}^N_{Y\|X}=\sup_{f_{xy,N}:\mathcal{X}^N\times\mathcal{Y}^N\mapsto\mathbb{R}}\mathbb{E}\left[f_{xy,N}(Y^0_{-(N-1)},X^0_{-(N-1)})\right]-\log\mathbb{E}\left[e^{f_{xy,N}(\widetilde{Y},Y^{-1}_{-(N-1)},X^{-1}_{-(N-1)})}\right], \tag{5}$$

$$\mathsf{D}^N_{Y}=\sup_{f_{y,N}:\mathcal{Y}^N\mapsto\mathbb{R}}\mathbb{E}\left[f_{y,N}(Y^0_{-(N-1)})\right]-\log\mathbb{E}\left[e^{f_{y,N}(\widetilde{Y},Y^{-1}_{-(N-1)})}\right], \tag{6}$$

其中，式（5）和式（6）中达到上确界的 DV 势函数分别为

$$f^\star_{y,N}:=\log\frac{p_{Y_0|Y^{-1}_{-N}}}{p_{\widetilde{Y}}},\qquad f^\star_{xy,N}:=\log\frac{p_{Y_0|X^0_{-N}Y^{-1}_{-N}}}{p_{\widetilde{Y}}}. \tag{7}$$

[3]此公式假设满足所谓的信息稳定性（参见 [27]）。

随后，我们使用循环神经网络近似最优 DV 势函数（7），并通过对联合分布的输入输出序列求样本均值来估计期望。我们首先定义循环神经网络类 [29]。

定义 2（循环神经网络函数类）固定 $k, d_i, d_o \in \mathbb{N}$。具有 k 个神经元、输入输出维度为（d_i, d_o）的循环神经网络类 $\mathcal{G}^{(d_i,d_o,k)}{}_{\mathsf{rnn}}$ 是具有以下结构的离散时间非线性系统集合:

$$\begin{aligned} s_{t+1} &= -\alpha s_t + \mathrm{A}\sigma(s_t + \mathrm{B}x_t), \\ y_t &= \mathrm{C}s_t, \end{aligned}$$

其中，$s_t \in \mathbb{R}^k$、$x_t \in \mathbb{R}^{d_i}$ 和 $y_t \in \mathbb{R}^{d_o}$ 分别是状态、输入和输出（列）向量；$\mathrm{A} \in \mathbb{R}^{k\times k}$、$\mathrm{B} \in \mathbb{R}^{k\times d_i}$ 和 $\mathrm{C} \in \mathbb{R}^{d_o\times k}$ 是相应的权重矩阵；$\alpha \in (-1,1)$ 是用于控制状态衰减的固定常数；$\sigma(x) = \frac{1}{1+e^{-x}}$ 是 Sigmoid函数，逐分量作用于向量。

注意，$\mathcal{G}^{(d_i,d_o,k)}{}_{\mathsf{rnn}}$ 是一个参数化类，其（有限多个）参数属于某个参数空间 $\Theta \subset \mathbb{R}^d$，维度为适当的 d。当 k 固定时，我们可以互换使用显式形式 $g \in \mathcal{G}^{(d_i,d_o)}{}_k$ 或参数化形式 g_θ 来表示上述类中的函数，其中 $\theta \in \Theta$。采用这一符号，DINE目标函数表示为 [22]

$$\widehat{\mathsf{I}}_{\mathsf{DI}}(D_n, \theta_y, \theta_{xy}) := \sup_{\theta_{xy}\in\Theta_{xy}} \widehat{\mathsf{D}}_{Y\|X}(D_n, \theta_{xy}) - \sup_{\theta_y\in\Theta_y} \widehat{\mathsf{D}}_y(D_n, \theta_y), \tag{8}$$

其中，$\theta_y \in \Theta_y$ 和 $\theta_{xy} \in \Theta_{xy}$ 分别是循环神经网络 $g_{\theta_y} \in \mathcal{G}^{(d_y,1,k)}{}_{\mathsf{rnn}}$ 和 $g_{\theta_{xy}} \in \mathcal{G}^{(d_y+d_x,1,k)}{}_{\mathsf{rnn}}$ 的参数，KL散度估计器定义为

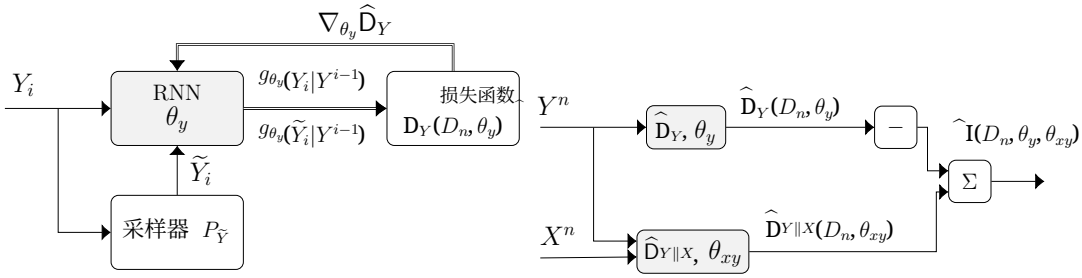
$$\widehat{\mathsf{D}}_Y(D_n, \theta_y) := \frac{1}{n}\sum_{i=1}^n g_{\theta_y}\left(Y^i\right) - \log\left(\frac{1}{n}\sum_{i=1}^n e^{g_{\theta_y}\left(\widetilde{Y}_i, Y^{i-1}\right)}\right), \tag{9a}$$

$$\widehat{\mathsf{D}}_{Y\|X}(D_n, g_{\theta_{xy}}) := \frac{1}{n}\sum_{i=1}^n g_{\theta_{xy}}\left(Y^i, X^i\right) - \log\left(\frac{1}{n}\sum_{i=1}^n e^{g_{\theta_{xy}}\left(\widetilde{Y}_i, Y^{i-1}, X^i\right)}\right), \tag{9b}$$

$\widetilde{}$和 Y^n 是从 Unif($\mathcal{Y}$) 中抽取的独立同分布序列。现在，DINE 通过在相应的参数空间上优化两个 KL 散度估计器得到:

$$\begin{aligned} \widehat{\mathsf{I}}_{\mathsf{DI}}(D_n) &:= \sup_{\theta_{xy}\in\Theta_{xy}} \inf_{\theta_y\in\Theta_y} \widehat{\mathsf{I}}_{\mathsf{DI}}(D_n, \theta_y, \theta_{xy}) \\ &= \sup_{\theta_{xy}\in\Theta_{xy}} \widehat{\mathsf{D}}_{Y\|X}(D_n, \theta_{xy}) - \sup_{\theta_y\in\Theta_Y} \widehat{\mathsf{D}}_y(D_n, \theta_y), \end{aligned}$$

DINE架构如图 1 所示。有关DINE的形式一致性保证以及实现细节，请参阅 [22]。



(a) $\widehat{\mathsf{D}}_Y$ 模型。 (b) DINE架构。

图1：DINE模型。图(a)展示了单个KL散度估计器的实现，图(b)展示了完整的DINE系统。

2.4 马尔可夫决策过程

马尔可夫决策过程是用于随机系统中序贯决策的离散时间随机控制过程 [30]。马尔可夫决策过程由元组（$\mathcal{Z}$、$\mathcal{U}$、$\mathcal{W}$、$P_{W|Z}$、υ、f、r）描述，其中 $\mathcal{Z}$和 $\mathcal{U}$分别是状态空间和动作空间，$W \sim P_{W|Z}$，$\upsilon(\cdot|z,\ u) \in \mathcal{P}(\mathcal{W})$ 是在给定（z、u）$\in \mathcal{Z}\times\mathcal{U}$、$r$时的扰动：$\mathcal{U}\times\mathcal{Z}\to\mathbb{R}$是即时奖励，函数 f：$\mathcal{Z}\times\mathcal{U}\times\mathcal{W}\to\mathcal{Z}$描述状态演化，即 $z_{t+1}=f(z_t,\ u_t,\ w_t)$。动作由随机策略 $\boldsymbol{\pi}=(\pi_t)_{t\in\mathbb{N}}$决定，其中每个 π_t 都是给定 Z_t时 U_t 的条件分布，即如果 $Z_t=z$，则对每个 $t\in\mathbb{N}$都有 $U_t\sim\pi_t(\cdot|z)\in\Delta_{\mathcal{U}}$。

我们考虑一个无限时域平均奖励马尔可夫决策过程，其目标定义为

$$\rho(\boldsymbol{\pi}) := \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}_{\boldsymbol{\pi}}\left[r(U_t,Z_t)\right], \tag{10}$$

其中，下标 $\boldsymbol{\pi}$ 强调它诱导了该序列分布。马尔可夫决策过程智能体的目标是找到一个策略 $\boldsymbol{\pi}$，使 $\rho(\boldsymbol{\pi})$最大化。虽然先验而言，最优 $\boldsymbol{\pi}$ 可以是任意的，但事实证明，对于任意无限时域马尔可夫决策过程，若 $|\mathcal{Z}|<\infty$，则极大点$_{\boldsymbol{\pi}}$ $\rho(\boldsymbol{\pi})$中包含一个平稳策略 [30], ，即存在某个条件分布π，使得对所有 $z\in\mathcal{Z}$ 和 $t\in\mathbb{N}$，均有 $\pi_t(\cdot|z)=\pi(\cdot|z)$。

3 定向信息优化

我们提出了一种利用输入过程概率质量函数的基于RNN的生成模型，在离散输入空间上优化DINE的新方法。为得到一个易于处理的优化目标，我们将定向信息率优化问题表述为马尔可夫决策过程，并结合函数逼近结果和蒙特卡洛方法，运用策略梯度定理[23] 。

下文中，$\mathbb{X}$ 和 $\mathbb{Y}$ 表示联合平稳的离散时间随机过程，其样本分别取值于 $\mathcal{X}$ 和 $\mathcal{Y}$。尽管本方法普遍适用，我们在通信信道的背景下介绍该方法，其中 $\mathcal{X}$ 和 $\mathcal{Y}$ 分别表示信道输入空间和输出空间。我们假设输入字母表$|\mathcal{X}|=k<\infty$ 的大小已知。为便于说明，我们聚焦于 $\mathcal{Y}$ 也

表1：DI速率优化MDP表述

MDP	DI优化
状态 Z_t	X_{-t}^{-1}, Y_{-t}^{-1}
动作 U_t	X_0
扰动 W_t	Y_0
奖励 $r(U_t, Z_t)$	式（12）

有限且信道由因果条件概率质量函数$p_{Y^n\|X^n}$描述的情形。不过，我们的推导与输出字母表无关，也可轻松扩展至具有连续输出的信道。

3.1 互信息速率优化问题

我们构建定向信息率优化问题（见第2.2节），并按照下文所述，使用深度生成模型对输入过程概率质量函数进行建模。该生成模型接收来自 $\mathcal{X} \times \mathcal{Y}$ 的一个输入输出对以及一个单纯形向量（用于对当前输入概率质量函数建模），并输出一个新的单纯形向量（更新后的输入概率质量函数）。与参数向量 $\phi \in \Phi \subset \mathbb{R}^d$ 对应的概率质量函数生成器记为 h_ϕ：$\mathcal{X} \times \mathcal{Y} \times \Delta_k \to \Delta_k$。由于每个新的单纯形向量 p_t^ϕ 由参数和过去输入输出对序列（X^{t-1}, Y^{t-1}）确定，因此我们将 t的第 h_ϕ 个输出视为给定过去序列时 X_t的条件概率质量函数模型:

$$p_t^\phi = p^\phi(\cdot|X^{t-1}, Y^{t-1}) := h_\phi(X_{t-1}, Y_{t-1}, p_{t-1}^\phi), \qquad t \geq 1,$$

其中，对于每个 $s \leq t-1$，$(X_s, Y_s) \sim p_s^\phi p_{Y_s|X^sY^{s-1}}$。

现在的目标是在由h_ϕ、 $\phi \in \Phi$建模的所有输入分布上优化DI速率，即求解

$$\sup_{\phi \in \Phi} \mathsf{I}_\phi(\mathbb{X} \to \mathbb{Y}), \tag{11}$$

其中，下标 ϕ 表示 DI 速率表达式中每个$\mathsf{I}(X^n \to Y^n)$ 的底层分布为$n \in \mathbb{N}$, $\prod_{t=1}^n p_t^\phi p_{Y_t|X^tY^{t-1}}$。注意，由于联合分布的平稳性，$\mathsf{I}_\phi(\mathbb{X} \to \mathbb{Y})$ 存在；当 Φ 为紧致集时，（11）中的上确界可以达到。为求解（11），我们需要找到定向互信息率梯度 $\nabla_\phi \mathsf{I}_\phi(\mathbb{X} \to \mathbb{Y})$ 的可处理表达式。下一节将互信息率优化重新表述为马尔可夫决策过程，并运用策略梯度定理，结合基于 DINE 的马尔可夫决策过程奖励近似，得到一个梯度与上述梯度一致的目标函数。最后，该节将介绍一种深度强化学习策略优化方法，用于计算（11）。

3.2 MDP 表述

回顾一下，MDP 由元组 $(\mathcal{Z}, \mathcal{U}, \mathcal{W}, P_{W|Z,\ U}, f, r)$ 给出。根据模型的平稳性，我们可以在每个时间步应用反向时间移位算子，使最新的时间步始终为 $t = 0$。为了得到 DI 速率优化的 MDP 表述，我们

将过去信道输入和输出的累积量作为状态，即 $Z_t = (X_{-t}^{-1}, Y_{-t}^{-1})$。[4]我们将信道输入生成器视为一个智能体，其在每个时间步采取的动作 $U_t = X_0$ 由参数化策略 $\pi_\phi(\cdot|Z_t) = p_t^\phi(\cdot|X_{-t}^{-1}, Y_{-t}^{-1})$ 抽取。扰动是信道输出，其服从条件概率质量函数 $p_{Y_0|Y_{-t}^{-1}X_{-t}^0}$，即时奖励则由条件期望给出

$$r(U_t, Z_t) = \mathbb{E}\left[\log\left(\frac{p_{Y_0|Y_{-t}^{-1},X_{-t}^0}(Y_0|Y_{-t}^{-1}, X_{-t}^0)}{p_{Y_0|Y_{-t}^{-1}}(Y_0|Y_{-t}^{-1})}\right)\middle| X_{-t}^0, Y_{-t}^{-1}\right], \tag{12}$$

对于这些情况，我们有 $\mathbb{E}[r(U_t, Z_t)] = \mathsf{I}^\phi(X_{-t}^0; Y_0|Y_{-t}^{-1})$。对于前馈通信，MDP 表述保持不变，但优化仅限于不依赖过去信道输出的策略。该表述总结于表 1，并由此得到所需的 MDP 刻画（证明见第 8.1 节）。

定理1（MDP 表述）定向信息率优化问题 (11) 是一个无限时域平均奖励马尔可夫决策过程，其目标为 $\rho(\pi_\phi) = \mathsf{I}^\phi(\mathbb{X} \to \mathbb{Y})$。

注释1（与现有 MDP 的关系）此前曾使用 MDP 表述来计算某些单叶有限状态信道的容量优化问题的反馈容量 *[10,11,15]*，并假设信道模型已知。在这些表述中，马尔可夫决策过程状态空间是 $\Delta_{|\mathcal{S}|}$ 的量化版本，且 $\rho(\pi)$ 是一个单字母表达式，可通过动态规划算法进行优化。文献 *[31]* 的作者将单叶表述加以推广，但得到的目标函数通常难以处理。另一方面，我们的表述可以看作适用于所有具有平稳联合分布的信道的统一 MDP。与 *[31]*, 一样，我们的 MDP 无法用传统方法计算，因此需要借助近似和估计技术提出新思路。

注释2（最优解的存在性）对于状态空间和动作空间有限的马尔可夫决策过程，能够保证存在最优策略 *[30]*。然而，当状态空间变为无限时，除非马尔可夫决策过程满足关于其遍历性的某些条件，否则无法保证存在最优策略（参见 *[32–35]*）。此前使用马尔可夫决策过程最大化互信息的方法便属于这种情况，其马尔可夫决策过程状态空间是一个概率单纯形 *[10, 11, 15, 16, 36]*。

3.3 策略梯度定理

我们利用 MDP 表述，借助强化学习技术，得到 $\nabla_\phi \rho(\pi_\phi)$ 的一个易于处理的表达式。此方法尤其适用于此处，因为我们将信道视为黑箱，可根据输入序列对其进行采样，而强化学习为求解数据驱动的马尔可夫决策过程提供了强大的工具。我们引用策略梯度定理 [23, 定理 1], , 该定理能够将 $\nabla_\phi \rho(\pi_\phi)$ 表示为 $\nabla_\phi \pi_\phi$ 的函数，而后者通常更易于计算。

定理 2 （策略梯度）设 $(\mathcal{Z}, \mathcal{U}, \mathcal{W}, P_{W|Z,U}, f, r)$ 为一个无限时域平均

[4]为了确保所有 t 的状态空间相同，我们将状态 $Z_t = (X_{-t}^{-1}, Y_{-t}^{-1})$ 与无穷多个空符号连接起来，使得 $\mathcal{Z}$ 构成一个半无限序列空间。

奖励马尔可夫决策过程，其目标为 ρ，并考虑参数化的平稳策略 π_ϕ，其中 $\phi \in \Phi \subseteq \mathbb{R}^d$，对于某个 $d \in \mathbb{N}$。定义函数

$$\mathsf{Q}^{\pi_\phi}(u,z) := \sum_{t=1}^{\infty} \mathbb{E}\left[r(U_t,Z_t) - \rho(\pi_\phi) \middle| Z_0 = z, U_0 = u\right]. \tag{13}$$

那么，

$$\nabla_\phi \rho(\pi_\phi) = \sum_{z\in\mathcal{Z}} p_Z^{\pi_\phi}(z) \sum_{u\in\mathcal{U}} \nabla_\phi \pi_\phi(u,z) \mathsf{Q}^{\pi_\phi}(u,z), \tag{14}$$

其中，$p_Z^{\pi_\phi}$ 是马尔可夫决策过程状态序列的平稳分布。

状态-动作价值函数 Q^{π_ϕ} 式（13）量化了给定状态-动作对下未来奖励相对于 ρ（π_ϕ）的期望偏差。策略梯度定理通过用策略梯度 $\nabla_\phi \rho$（π_ϕ）和函数 $\nabla_\phi \pi_\phi$ 来表示，从而简化了 Q^{π_ϕ}。利用恒等式 $\partial_x \log f$（x）$= \frac{\partial_x f}{f}$（x）f（x），我们可以进一步表示式（14）

$$\begin{aligned}\nabla_\phi \rho(\pi_\phi) &= \sum_z p_Z^{\pi_\phi}(z) \sum_u \pi_\phi(u,z) \nabla_\phi \log\left(\pi_\phi(u,z)\right) \mathsf{Q}^{\pi_\phi}(u,z) \\ &= \mathbb{E}\left[\nabla_\phi \log\left(\pi_\phi(U,Z)\right) \mathsf{Q}^{\pi_\phi}(U,Z)\right]. \end{aligned} \tag{15}$$

接下来　使用 DINE 近似 Q^{π_ϕ}，并得到式（15　　）.

3.4 通过 DINE 进行近似

为了通过式（15）的右端（RHS）计算所需的梯度，必须计算函数 Q^{π_ϕ}。然而，这需要计算马尔可夫决策过程的奖励（12），而这又需要了解信道模型。即使已知信道，由于马尔可夫决策过程状态空间规模过大，也无法高效地应用传统表格算法。我们通过使用 DINE [22]，仅基于信道样本来近似 Q^{π_ϕ}，从而解决这一问题。

固定概率质量函数生成器参数 $\phi \in \Phi$，并考虑从该输入概率质量函数和信道 $D_n = (X^n, Y^n) \sim \prod_{t=1}^n p_t^\phi p_{Y_t|Y^{t-1}X^t}$抽取的数据集。首先，我们将 DINE $\widehat{\mathsf{I}}_{\mathsf{DI}}(D_n)$ 作为真实 DI速率 $\rho(\pi_\phi)$ 的强一致估计值，参见 [22, 定理 2]。然后，使用训练后的 DINE 循环神经网络（g_{θ_y}, $g_{\theta_{xy}}$）近似 Q^{π_ϕ} 定义中的函数 $r(U_t, Z_t)$，如下所示。我们考虑式（5）和（6）中所导出KL散度的 DV表示。减去其达到上确界的 DV势函数（7），即可得到似然比 $f^\star_{xy,N} - f^\star_{y,N} = \log\left(\frac{p_{Y_0|X_{-N}^0 Y_{-N}^{-1}}}{p_{Y_0|Y_{-N}^{-1}}}\right)$，其中

$$\mathbb{E}[r(U_N,Z_N)] = \mathbb{E}\left[f^\star_{xy,N}(X_{-N}^0, Y_{-N}^0) - f^\star_{y,N}(Y_{-N}^0)\right].$$

随着 DINE循环神经网络 $(g_{\theta_{xy}}, g_{\theta_y})$ 经过优化，以达到与 (5) 和 (6) 相对应的经验 DV 形式的上确界，我们定义 $\widehat{r}_\theta := g_{\theta_{xy}} - g_{\theta_y}$，并将其作为 r 的替代指标。这一构造假设 $\phi \in \Phi$ 固定不变，且 g_{θ_y} 和 $g_{\theta_{xy}}$ 已针对诱导的联合分布 $\prod_{t=1}^n p_t^\phi p_{Y_t|Y^{t-1}X^t}$进行了优化。我们将在后文进一步讨论这一假设

第 4 节将提出联合优化算法。

在近似 Q^{π_ϕ} 的最后一步中，我们观察到

$$\lim_{t\to\infty} \mathbb{E}\left[r(U_t, Z_t)\right] = \lim_{t\to\infty} \mathsf{I}_\phi(X_{-t}^0; Y_0|Y_{-t}^{-1}) = \mathsf{I}_\phi(\mathbb{X}\to\mathbb{Y}) = \rho(\pi),$$

这意味着，当 t 增大时，差值 $r(U_t, Z_t) - \rho(\pi_\phi)$ 会变得可以忽略不计。这为在某个 $T < \infty$ 处截断定义 Q^{π_ϕ} 的无穷和提供了依据；结合上述步骤，可得近似

$$\widehat{\mathsf{Q}}_{\theta,t}(D_n) := \sum_{i=t}^{t+T-1} \widehat{r}_\theta(Y^i, X^i) - \widehat{\mathsf{I}}(D_n, \theta).$$

有了这一点，我们通过对一条较长轨迹 $(p_t^\phi, X_t, Y_t)_{t=1}^n$进行蒙特卡洛评估，以此替代外层期望，从而估计 (15) 的右侧，得到

$$\nabla_\phi \underbrace{\frac{1}{n-T}\sum_{t=1}^{n-T} \log\left(p_t^\phi(X_t)\right) \widehat{\mathsf{Q}}_{\theta,t}(D_n)}_{=:\widehat{\mathsf{J}}_\theta(D_n,\phi)} \tag{16}$$

作为 $\nabla_\phi \rho(\pi_\phi)$ 的最终近似。仅基于来自输入生成模型和信道的样本，该目标关于 ϕ 易于求导，正如所需。因此，$\widehat{\quad}$ $\widehat{\quad}$DI 优化方案交替优化 $\mathsf{I}(D_n)$ 和 $\mathsf{J}_\theta(D_n, \phi)$，即分别交替进行 $\widehat{r}_\theta$ 的近似改进和策略优化。

备注 3（性能保证）尽管 DINE 是 DI*[22],* 的一致估计器，但要为我们整体方法提供形式化保证，需要从理论上分析基于循环神经网络的策略优化与蒙特卡洛方案的结合，而目前尚无此类分析 *[37]*。尽管如此，在第 *5* 节中，我们通过实验证明，我们的方法在所有考虑的示例中都表现良好。

3.5 互信息估计优化器

当信道无记忆时，优化问题 (11) 可简化为在$\mathsf{I}\phi(X;Y)$ 上进行最大化，其中输入生成模型属于单纯形 Δ_k。在这个特例中，我们可以利用无记忆结构，采用更简单的模型。首先，h_ϕ 不再需要依赖过去的信道输入—输出对，因此可以取为$p^\phi = \left(\sigma_{\mathsf{sm}}^1(\phi)\ldots, \sigma_{\mathsf{sm}}^m(\phi)\right)$，其中

$$\sigma_{\mathsf{sm}}^i(\phi) := \frac{\exp(\phi_i)}{\sum_{j=1}^m \exp(\phi_j)}, \qquad i = 1, \ldots, m \tag{17}$$

是 softmax 函数的第 i 个输出。

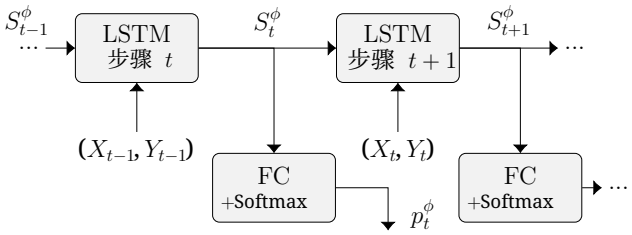


图2：用于反馈容量的概率质量函数模型展开。在第 t 步中，根据（S^{ϕ}_{t-1}、X_t、Y_t）计算 S^{ϕ}_{t}，然后将其用于计算 S^{ϕ}_{t+1} 和 p^{ϕ}_{t}。

其次，我们用互信息神经估计器（MINE）替换 DINE [38]:

$$\widehat{\mathsf{I}}_{\mathsf{MI}}(D_n) := \sup_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} g_\theta(X_i, Y_i) - \log\left(\frac{1}{n}\sum_{t=1}^{n} e^{g_\theta(X_t, \bar{Y}_t)}\right), \tag{18}$$

其中 g_θ 是一个前馈神经网络，$D_n = (X^n、Y^n) \overset{i.i.d.}{\sim} p^{\phi} p_{Y|X}$ 以及 $\bar{Y}_t$ 是与 X_t 相互独立的负样本，适用于 $t = 1, \ldots n$。互信息神经估计器在前馈神经网络上进行优化，这类网络更为简单，通常具有更好的收敛特性。此外，互信息神经估计器对真实互信息 [22, 给出下界 5]（DINE 不保证这一点），并且遵循非渐近误差界 [39, 40]。在第 7 节中，我们将展示基于互信息的优化方案，以学习峰值功率受限 AWGN（PP-AWGN）中的星座概率整形。

由此得到的求导目标形式如下:

$$\widehat{\mathsf{J}}^{\mathsf{MI}}_{\theta}(D_n, \phi) := \frac{1}{n}\sum_{t=1}^{n} \log\left(p^{\phi}(X_t)\right)\left(g_\theta(X_t, Y_t) - \widehat{\mathsf{I}}_{\mathsf{MI}}(D_n)\right), \tag{19}$$

其梯度可简化如下（证明见第 8.2 节）。

引理 1 式 (19) 关于 ϕ 的梯度为

$$\nabla_\phi \widehat{\mathsf{J}}^{\mathsf{MI}}_{\theta}(D_n, \phi) = \frac{1}{n}\sum_{t=1}^{n} \left(e_{X_t} - p^{\phi}\right)\left(g_\theta(X_t, Y_t) - \widehat{\mathsf{I}}_{\mathsf{MI}}(D_n)\right), \tag{20}$$

其中 $e_{X_t} = \left(\mathbb{I}_{\{X_t=1\}} \cdots \mathbb{I}_{\{X_t=m\}}\right)^{\mathsf{T}}$ 是一个 m维标准基向量，其第 X_t个分量为 1。

4 实现与算法

4.1 概率质量函数生成器

回顾一下，概率质量函数生成器计算映射 h_ϕ：$\mathcal{X} \times \mathcal{Y} \times \Delta_k \to \Delta_k$，其输出按照 $p^{\phi}_{t} = h_\phi(X_{t-1}, Y_{t-1}, p^{\phi}_{t-1})$ 演变。我们使用长短期记忆网络

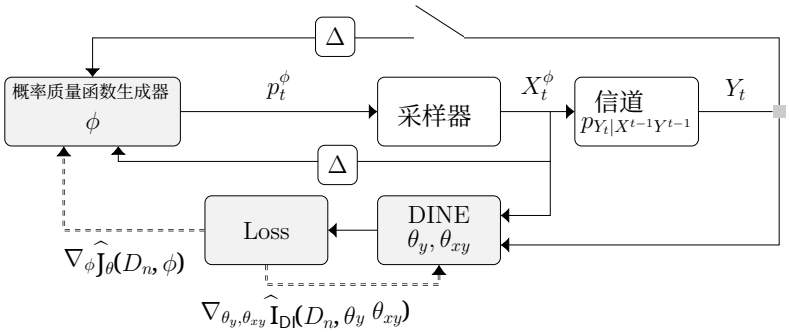


图3：完整的估计优化模型。虚线箭头表示梯度传播，实心方块表示参数化模型。

（LSTM）网络实现 。LSTM 是一种循环神经网络，它利用门控机制在多个时间步中选择性地保留、输入、输出和遗忘信息，从而能够对序列数据中的长期依赖关系建模（有关 LSTM 的更多背景信息，请参见 [41] ）。我们将 LSTM 网络与额外的全连接网络堆叠，以增强模型的表达能力。 h_ϕ 的输出层由m维 Softmax 激活（17）构成。分别用 g_1^ϕ 和 g_2^ϕ 表示 LSTM 和 FC 映射，则概率质量函数生成器的输出演变可表示为

$$S_t^\phi = g_1^\phi(X_{t-1}, Y_{t-1}, S_{t-1}^\phi), \quad p_t^\phi = \sigma_{\mathrm{sm}}\big(g_2^\phi(S_t^\phi)\big), \tag{21}$$

其中 S_t^ϕ 是时刻 t的长短期记忆网络内部状态。当考虑前馈容量时，会从（21）中省略 Y_{t-1}。h_ϕ 的架构如图 2 所示。

4.2 组合系统

如图 3所示，组合系统由概率质量函数生成器和 DINE模型组成。因此，我们构建了一个联合估计优化过程，通过交替最大化 $\widehat{\mathsf{I}}_{\mathsf{DI}}(D_n, \theta_y, \theta_{xy})$ 和 $\mathsf{J}_\theta(D_n, \phi)$。每次迭代中，选择一个模型进行参数更新，同时固定另一个模型的参数。优化 DINE可提高固定输入PMF模型 Q^{π_ϕ} 时的近似精度 h_ϕ；而优化 h_ϕ 则会提高互信息，其大小由当前 DINE 模型量化。

每次迭代中，我们执行以下步骤：首先，依次计算概率质量函数、从中采样，并将采样得到的输入通过信道传播，从而计算 $p^{\phi,\ n}$ 和数据集 D_n。然后，将 D_n 输入 DINE，并计算损失函数。我们使用随机梯度上升更新模型参数；对于 θ 更新，计算 $\nabla_{\theta_y,\ \theta_{xy}} \mathsf{I}_{\mathsf{DI}}(D_n, \widehat{\theta}_y,\ \theta_{xy})$；对于 ϕ 更新，计算 $\nabla_\phi \mathsf{J}_\theta(D_n,\ \phi)$。重复这些步骤，直至满足收敛准则或完成预定次数的更新。最后，在一长串信道输入和输出上评估 DINE目标函数，以获得优化后互信息的数值估计值。完整步骤见算法 1 。

所提出的联合优化方法涉及一项潜在假设；更新概率质量函数

生成器需要准确估计相对于当前 ϕ 值所诱导的联合分布的 Q^{π_ϕ}。因此，我们优先训练 DINE模型，并且每次更新概率质量函数生成器参数 ϕ 时，都对 DINE参数（θ_y、θ_{xy}）进行多次更新。

算法1 离散字母表DI优化与估计

输入： 离散信道、反馈指示器

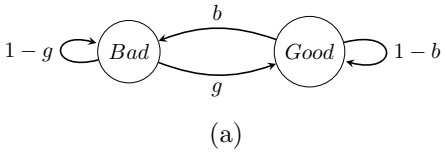
输出： $\widehat{\mathrm{I}}_{\mathrm{DI}}(D_n, h_\phi)$，优化后的 h_ϕ。

使用参数 θ_y、θ_{xy}和 ϕ初始化g_{θ_y}、$g_{\theta_{xy}}$、h_ϕ，并设置学习率 γ。
如果存在反馈指示器，则
  向 h_ϕ 添加反馈链路
重复
  计算 $(D_n, p^{\phi,\ n})$
  如果是训练DINE，则
    计算 $\widehat{\mathrm{D}}_Y(D_n, \theta_y)$，根据（$\mathrm{D}^{Y\|X}(D_n, \theta_{xy})$9）
    更新 DINE参数：
    $\theta_y \leftarrow \theta_y + \gamma \nabla_{\theta_y} \widehat{\mathrm{D}}_Y(D_n, \theta_y)$
    $\theta_{xy} \leftarrow \theta_{xy} + \gamma \nabla_{\theta_{xy}} \widehat{\mathrm{D}}^{Y\|X}(D_n, \theta_{xy})$
  否则（训练概率质量函数生成器）
    计算 $\widehat{\mathrm{J}}_\theta(D_n, \phi)$ 根据 (16)
    更新概率质量函数生成器参数：
    $\phi \leftarrow \phi + \gamma \nabla_\phi \widehat{\mathrm{J}}_\theta(D_n, \phi)$
直至收敛
蒙特卡洛评估的 $\widehat{\mathrm{I}}_{\mathrm{DI}}(D_n)$
返回 $\widehat{\mathrm{I}}_{\mathrm{DI}}(D_n)$ 和 h_ϕ

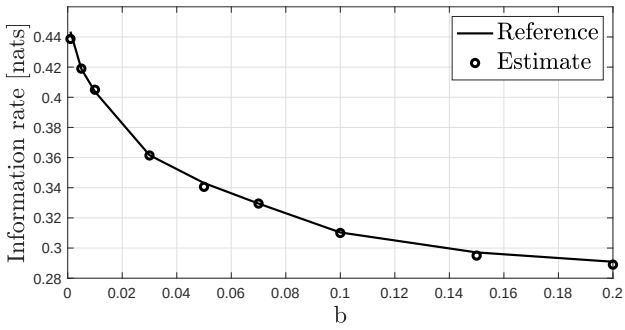
5 应用 1：容量估计

我们采用算法 1 来估计各种有记忆信道的容量。性能通过将优化后的 DI 估计值与已知容量解和/或界进行比较来衡量。此外，我们将学习得到的概率质量函数结构与已知的容量逼近编码方案进行比较。与本文提出的方法不同，所有纳入比较的参考方法都需要完整的信道律知识。实现代码可在 GitHub 上找到。

注释 4（容量优化问题）所提出的方法旨在计算一组输入分布下 DI 速率的上确界。然而，容量表达式(2) 采用了极限与上确界的相反顺序，即对经过优化的归一化DI序列取极限。已知对于有限状态信道*[42]*，（FSC）而言，这两种顺序可以互换；有限状态信道涵盖了大多数有记忆信道模型。



(a)



(b)

图4：GE信道。图(a)展示信道状态马尔可夫链。从“良好状态”到“恶劣状态”及反向的转移分布分别为 Ber(b) 和 Ber(g)；(b)展示估计容量随 b（其中 $g = 3b$）的变化，并与基于 [44] 得到的估计值进行比较。

5.0.1 吉尔伯特-埃利奥特信道

吉尔伯特-埃利奥特（GE）信道是一种时变二元对称信道（BSC），其翻转参数由潜在马尔可夫状态决定，该状态按照图4a中的状态图演变。在“良好状态”下，翻转概率较低，因此信道性能更好。虽然很容易证明，最优输入是独立同分布的 Ber(0.5) 过程，但 GE 信道容量由一个极限表达式确定，而非闭式公式 [43]。

我们应用所提出的方案来估计 GE信道的容量，并将结果与根据长输入-输出序列 H 得到的 $\mathbb{Y}|\mathbb{X}$ 的一致估计值：$= \lim_{n\to\infty} \frac{1}{n} H(Y^n|X^n)$ 进行比较。该方法假设 [44] 中的元素服从容量实现分布，即 $\mathbb{X}$ Ber(0.5)。如图 4b 所示，我们的方法在各种信道状态转移值下均能达到容量。

5.0.2 伊辛信道

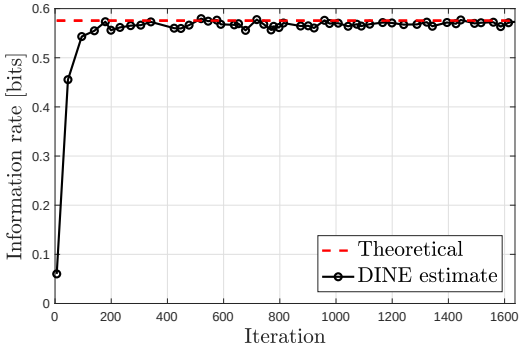
二元伊辛信道是一种单叶有限状态信道，其演化遵循[5]

$$Y_t = \begin{cases} \mathsf{Z}_{1/2}(X_t), & \text{if } S_{t-1} = 0 \\ \mathsf{S}_{1/2}(X_t), & \text{otherwise} \end{cases}, \qquad S_t = X_{t-1}, \tag{22}$$

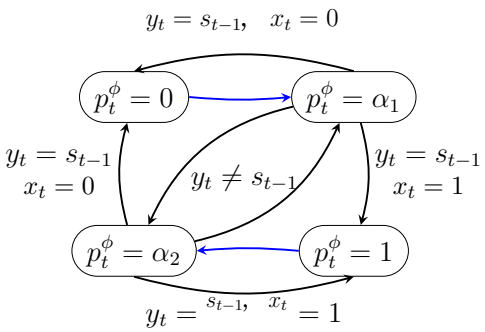
其中 $\mathsf{Z}^{1/2}$ 和 $\mathsf{S}^{1/2}$ 分别表示概率为 $1/2$的 Z 信道和 S信道。 [15]使用动态规划算法在量化状态空间上计算伊辛信道的容量。我们通过算法 1估计容量；经过相对较少的迭代次数后，该算法便收敛于真实容量；参见图 5a。

此外 我们通过分析学习到的概率质量函数的结构来评估我们的方法。我们

[5]当信 道模型已知时，由于信道是是单叶的。



(a)



(b)

图5：伊辛信道。图（a）展示了带反馈的伊辛信道上的 DINE 损失收敛情况，图（b）展示了针对（α_1，α_2）=（0.456，0.570）学习得到的概率质量函数模型。蓝色箭头表示对于（x_t，y_t，s_t）的任意取值都会发生的转移。

构造一条较长的轨迹（$p^{\phi,\,n}$，x^n，s^n，y^n），并对 k进行 K 均值聚类，其中 $p^{\phi,\,n}$取 $k = 4$。然后，我们根据过去的（p_t^ϕ，$p^{\phi,\,t-1}$，x^{t-1}，s^{t-1}，y^{t-1}）考察聚类后的 p_t^ϕ 的演化。在此情形下，$p_t^\phi \in [0,1]$ 被视为伯努利分布的参数。图 p_t^ϕ 展示了5b。我们注意到，联合估计优化过程产生了非平凡的输入概率质量函数演化，其结构与 [15]中提出的解析容量达成编码方案一致。

5.0.3 陷阱门信道

陷阱门信道是一个单叶有限状态信道，其状态按照以下方式演化：$S_t = S_{t-1} \oplus X_t \oplus Y_t$，其中 $\oplus$ 表示二进制异或运算，其输出由（22）给出。我们通过算法 1估计该信道的前馈容量和反馈容量。对于反馈容量，该算法从 [10], 收敛到解析解，如图 6a所示。

陷阱门信道的前馈容量是一个未解问题。在[45] 和 [46]中分别给出了容量值的上界和下界；前者采用了延迟反馈容量的界，而后者分别采用了 Blahut-Arimoto型算法。图 6b展示了已知界之间的 DINE收敛情况。具体而言，这为陷阱门信道的前馈容量提供了新的、更精确的估计。对多次运行结果取平均后，我们得到的数值为 $\widehat{C}^{\mathsf{FF}}_{\text{Trapdoor}} \approx 0.57246$。

5.0.4 NOST 和 POST 信道

作为最后一个例子，我们考虑噪声输出即状态（NOST）信道和前一输出即状态（POST）信道。信道输出由（22）给出，但参数 $p \in [0,1]$（而不是 $p = 1/2$）可以任意取值。NOST 信道状态根据 $S_t = \mathsf{Z}\eta(Y_t)$ 随机地演变，其中 $\eta \in [0,1]$；而 POST 信道是 $\eta = 0$ 的特例。

我们使用算法 1 来估计 POST 信道的前馈容量和 NOST 信道的反馈容量，分别考虑 p 和 η 的各种取值。

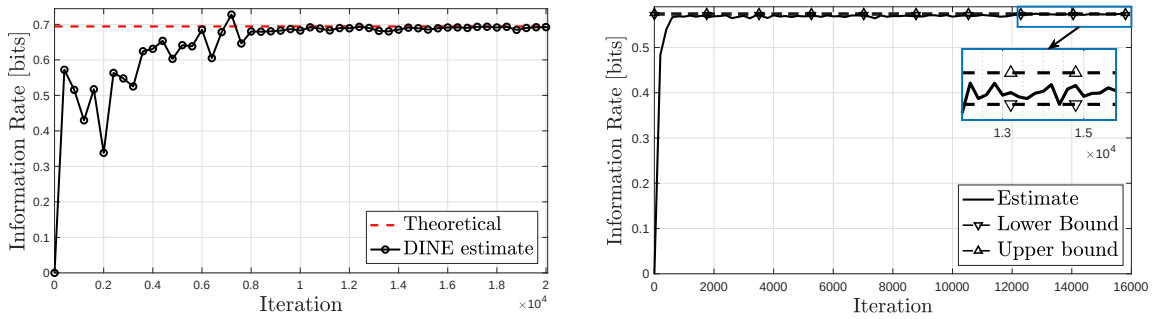


图6：图（a）和（b）分别展示了 DINE 损失函数收敛至反馈陷阱门信道容量和前馈陷阱门信道容量的过程。前馈容量估计值与 [19] 和 [46]，给出的上界和下界进行比较；反馈容量估计值则与 [10] 给出的解析解进行比较。

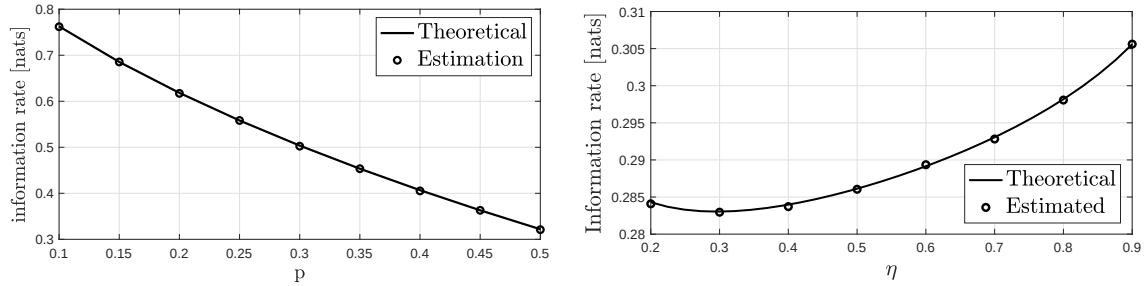


图7：图（a）展示了 POST 容量估计值与信道参数 p 的关系，并与解析容量值进行比较；图（b）展示了 NOST 信道的容量估计值与翻转概率 η 的关系，并与 [36] 中的方法进行比较。

图 7 将结果与 POST 信道的 [47] 以及 NOST 信道的 [36] 进行了比较。文献 [47] 的作者表明，POST(p) 信道的前馈容量和反馈容量均等于参数相同的 Z 信道容量。然而，前馈 POST 信道的容量实现分布通常可能具有无限记忆。这一事实，加上我们的容量估计值的准确性，证明了我们的输入分布模型具有很强的表达能力。

6 应用2：基于 Q图的反馈容量界

我们基于 [18,24] 中的工具，开发了一种计算单叶有限状态信道反馈容量上界和下界的方法。该方法利用算法 1 的输出，处理一个涉及结构化辅助随机变量的优化问题，该变量称为 Q-图。我们首先介绍 Q-图及其在计算

容量界。随后，我们指出，当前图搜索方法的复杂度高得难以承受，并提出一种 Q-图近似算法，该算法基于算法 1 中的优化 PMF 生成器。我们从估计的 Q-图和由此得到的容量界两方面展示了该方法的性能。所提出的方法将容量估计与计算其上界和下界的方法相结合。

6.1 背景： Q-图

对于有限信道输出字母表 $\mathcal{Y}$， Q-图是一个有向连通图，包含$|\mathcal{Q}_g|$ 个节点， $|\mathcal{Y}|$ 且每个节点都有$y \in \mathcal{Y}$条不同的出边，每条边都有唯一标记 Q。图上的节点转移由一个时不变的确定性函数 f_Q : $\mathcal{Q}_g \times \mathcal{Y} \to \mathcal{Q}_g$给出。 [18] 的作者 Q证明了，对于任意给定的 Q_g-图$\underline{\mathcal{L}}$，单叶有限状态信道的反馈容量可以通过在某个输入分布集合上求解相应的优化问题，得到上界和下界。将相应的界记为Q_g（ $\overline{\mathcal{L}}$）和Q_g（ [24, ），则有定理 2,3]:

$$\underline{\mathcal{L}}(Q_g) \leq \mathsf{C}_{\mathsf{FB}} \leq \overline{\mathcal{L}}(Q_g), \tag{23}$$

其中， $\underline{\mathcal{L}}(Q_g)$和 $\overline{\mathcal{L}}(Q_g)$均由$\sup^{P_{X|S,\ Q}\in\mathcal{P}} \mathrm{I}(X,S;Y|Q)$给出，但所用的分布集合 $\mathcal{P}$不同。对于 $\overline{\mathcal{L}}(Q_g)$，我们取 $\mathcal{P}=\mathcal{P}_Q$，即能够诱导唯一平稳联合过程$(S_i,Q_i)_{i\in\mathbb{N}}$的输入分布集合。对于$\underline{\mathcal{L}}$ (Q_g)，则考虑 $\mathcal{P}_Q$ 中满足某些马尔可夫准则的一个子集。实践中，给定图 Q_g后，可通过数值求解上述问题得到上界和下界（更多详情见 [18] ）。我们的目标是找到 Q_g ，使其最大化 $\underline{\mathcal{L}}(Q_g)$或最小化 $\overline{\mathcal{L}}(Q_g)$。[6]

6.2 现有方法及其复杂度

作者 [18] 提出了一种基于穷举搜索的枚举变体，用于寻找给定基数 Q下的最佳图，该方法称为图池化。该方法需要针对所有考虑的 Q图求解上界和下界优化问题。如 [18], 所述， $\mathcal{Q}_g$ 的基数界目前尚不清楚。因此，图池化的运行时间通常没有上界；如果对所有含 $|\mathcal{Q}_g|$ 个节点的图进行搜索后，仍未能得到反馈容量的紧界，我们就会继续搜索所有大小为 $|\mathcal{Q}_g|+1$的图。下一个引理表明，图池化方法必须考虑的 Q图的数量关于 $|\mathcal{Q}_g|$ 呈指数增长（证明见第 8.3 节）。

引理 2（图池化方法的复杂度下界）固定 $|\mathcal{Q}_g|$ ，并令 N_{GP}表示图池化方法中考虑的大小为 $|\mathcal{Q}_g|$的 Q图的数量。则有 $N_{\mathsf{GP}} \geq e^{|\mathcal{Q}_g|\text{对数}|\mathcal{Q}_g|}$。

引理 2 表明，寻找具有较大基数的良好状态 Q-图在计算上负担沉重，即便每个优化问题的复杂度都相对较低。为克服这一问题，

[6] $|\mathcal{Q}_g|<\infty$ 这意味着相应的极小点和极大点非空；不过，一般无法保证它们重合或有交集。

算法 2 Q-图结构和 C_{FB} 界的计算

输入：离散信道、优化的 PMF 生成器 h_ϕ 和 $|\mathcal{Q}_g|$

输出：$\widehat{C}_{\mathsf{LB}}, \widehat{C}_{\mathsf{UB}}$.

使用参数 ψ 和学习率 γ 初始化 g_ψ 。

步骤 1：训练 $\boldsymbol{g_\psi}$

重复

使用 h_ϕ 和信道计算 (S^n, Y^n)。

使用 g_ψ 计算 $\{q_t^\psi\}_{t=1}^n$

计算交叉熵损失（24）并更新 ψ:

$\psi \leftarrow \psi - \gamma \nabla_\psi \mathcal{L}_{\mathsf{CE}}(Y^n, S^n, \psi)$

直到收敛

步骤 2：获得界

生成序列 $(q^{\psi,n}, Y^n)$

对 $q^{\psi,n}$ 执行 k-均值聚类

根据 $(q^{\psi,n}, Y^n)$ 计算 M_Q

计算 $\widehat{C}_{\mathsf{LB}}$, 从 M_Q得到C_{UB}

返回 $\widehat{C}_{\mathsf{LB}}, \widehat{C}_{\mathsf{UB}}, M_Q$

我们接下来提出一种基于 RNN 的近似方法，用于寻找潜在最优的 Q-图。该方法使用算法 1的输出，其训练时间与 $|\mathcal{Q}_g|$无关。

6.3 Q-图近似方法

一种自然的 Q-图选择是 $Q_g = p_{S_t|Y^t}$，因为它是若干单叶有限状态信道解中的 MDP 状态过程 [10, 11, 15, 36]。对于这些有限状态信道，已有研究证明 $p_{S_t|Y^t}$ 的取值属于 $\Delta_{|\mathcal{S}|}$的有限子集，这限制了 Q-图的搜索空间。在无法获得此类界的情况下，我们的方法将生成 $p_{S_t|Y^t}$的量化代理。该过程首先通过监督学习方案，利用长短期记忆网络近似 $p_{S_t|Y^t}$ ，然后通过 k-均值从学习得到的近似中提取图结构。

6.3.1 映射近似

我们设计了一个基于RNN的生成模型 g_ψ： $\Delta_{|\mathcal{S}|} \times \mathcal{Y} \to \Delta_{|\mathcal{S}|}$ 其参数为 $\psi \in \mathbb{R}^d$。该模型接收一系列信道输出 Y^n，这些输出由经过优化的概率质量函数模型和信道生成，并递归计算出一个序列 $(q_t^\psi)_{t=1}^n$。模型以$q_0 = 0$为初始状态，并按以下方式演化

$$q_t^\psi = g_\psi(q_{t-1}^\psi, Y_t), \quad t = 1, \ldots, n.$$

通过最小化分类交叉熵来训练该模型，使其近似 $p_{S_t|Y^t}$ 的演化，分类交叉熵定义为

$$\mathcal{L}_{\mathsf{CE}}(Y^n, S^n, \psi) := -\sum_{t=1}^{n} e_{S_t}^{\mathsf{T}} \log q_t^{\psi} = -\sum_{t=1}^{n} \log q_{t,S_t}^{\psi}, \tag{24}$$

其中，e_{S_t} 是 S_t的独热编码，而 q_t^{ψ} ，s_t 是 q_t^{ψ} 的第 S_t个坐标。

6.3.2 轨迹分析

在获得$p_{S_t|Y^t}$的循环神经网络近似后，我们旨在提取底层图结构。为此，我们构造一个长序列（q^{ψ}、^{n}、Y^n），并对 k进行 K 均值聚类，[48] 将其聚类为$_{q^{\psi},\ ^n}$ 使每个簇代表一个近似的图节点。接下来，我们根据任意两个节点之间最频繁的转移来标记边[7]。图的转移信息存储在一个三维二进制邻接数据结构中，记为 M_Q，其中，当从节点 $M_Q(i、\ j、\ k) = 1$ 到节点 i 的边 j 存在且标签为 $y = k$时，

我们将对应于 Q的图代入 M_Q 中的优化问题 [18]，并计算$\underline{\mathcal{L}}(M_Q)$ 和 $\overline{\mathcal{L}}(M_Q)$。如果得到的界不够紧，我们可以针对更大的图节点数 k重复分析步骤，并使用相同的轨迹（q^{ψ}、^{n}、Y^n）。完整方案见算法2。需要强调的是，虽然算法1不需要访问信道状态序列，但所提出的界计算方法确实依赖这一信息。

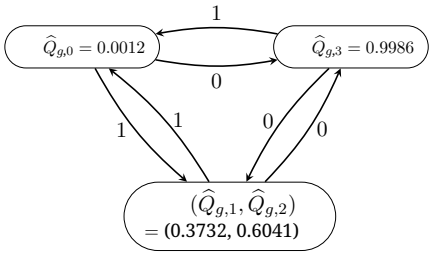
6.4 实证结果

我们展示了算法 2 在陷门信道和伊辛信道上的性能，因为对于这些示例，已知最优输入分布下 $p_{S_t|Y^t}$ 的结构。图8 和 9 将学习得到的 Q-图与 [15] 和 [10]中推导出的最优结构进行比较。估计的 $p_{S_t|Y^t}$ 与解析解所得结果一致，但与节点相关联的数值略有不同。尽管如此，这些调整后的数值不会影响界的计算；根据 [18], 中的方程（13）和（16），我们得到以下容量上界和下界：[18]

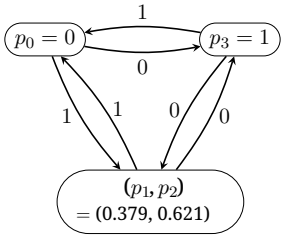
$$\text{Trapdoor:}\quad \widehat{C}_{\mathsf{UB}} - \widehat{C}_{\mathsf{LB}} \approx 3.7615 \cdot 10^{-08}, \quad \text{Ising:}\quad \widehat{C}_{\mathsf{UB}} - \widehat{C}_{\mathsf{LB}} \approx 2.4334 \cdot 10^{-07}$$

计算得到的界极为接近，充分证明了所提方法的有效性，同时也为算法 1提供了另一项实用的副产品。

[7] Q-图的广义版本可以考虑随机映射，因此也包括较少出现的转移。

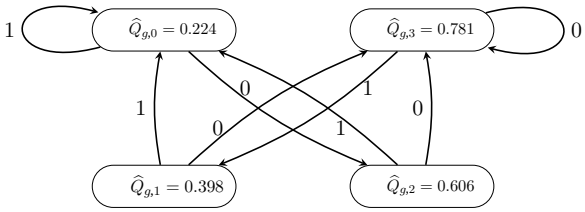


(a) 估计的 Q-图。

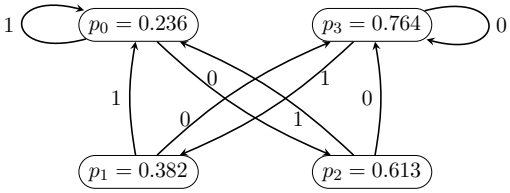


(b) 解析的马尔可夫决策过程状态。

图8：伊辛信道。估计的结构 (a) 与通过动态规划优化得到的马尔可夫决策过程状态转移 (b) 之间的比较。[15] (b)。



(a) 估计值 Q-图。



(b) 解析的马尔可夫决策过程状态。

图 9：陷阱门信道。估计结构 (a) 与通过动态规划优化得到的马尔可夫决策过程状态转移之间的比较 [10] (b)。

7 应用三：星座的概率整形

传统通信方案采用等概率星座 [25, 第 7]节。通过应用算法 1 进行互信息优化（见第 3.5节），我们提出了一种非均匀整形方案，并表明其通信速率优于等概率星座。考虑如下 PP-AWGN：

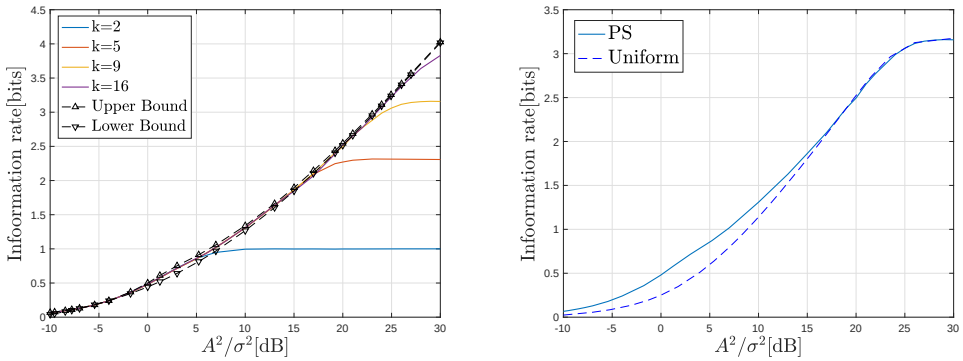
$$Y = X + Z, \qquad \text{s.t.} \quad |X| \le A, \quad P_X - a.s.,$$

其中 $A > 0$ 和 Z 是方差为 σ^2的中心高斯噪声。我们的目标是为 X 设计一个离散星座，使传输速率最大化。

7.1 实值加性高斯白噪声

Smith 证明，实值 PP-AWGN 信道的容量由一种支撑于 $[-A, A]$ 内的离散分布实现，其基数随 A^2/σ^2 [49] 增长。因此，我们考虑 $[-A, A]$ 内不同阶数的脉冲幅度调制星座。

图 10a 比较了互信息估计值与解析上 [50] 界和下界 [51], 在一系列 A^2/σ^2 取值下的结果。显然，互信息估计值收敛于两个界之间。请注意，互信息神经估计器本身构成真实互信息的下界 [22, 备注 5], 从而将我们的估计值作为峰值功率约束加性高斯白噪声信道容量的新数值下界。我们观察到，

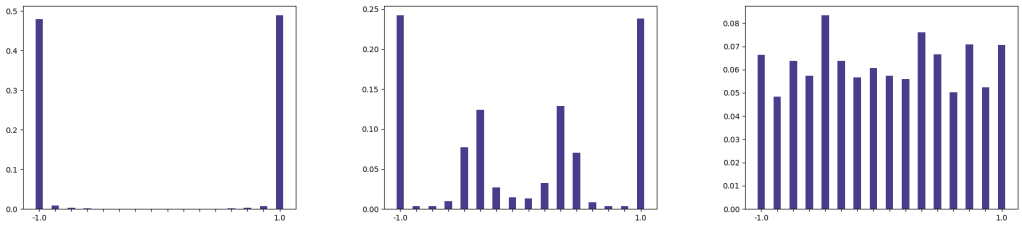


(a) (b)

图10：实值PP-AWGN的互信息估计值。图 (a) 展示了优化后的互信息与容量上界和下界的比较；图 (b) 展示了与 $k = 9$ 的均匀分布所产生的互信息的比较。

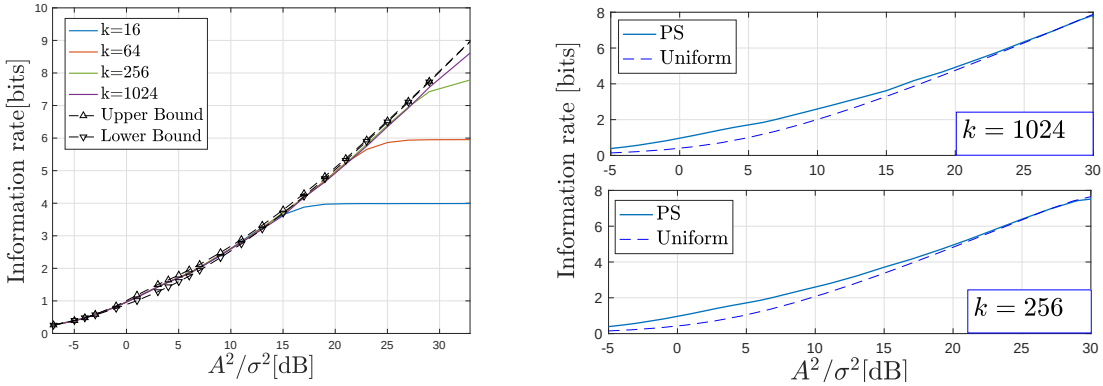
随着信噪比增大，所考虑的每种星座阶数下的信息速率都会饱和；这是由源熵上界 I$(X;\ Y)$ $\leq H(X)$ 导致的。图 10a 因此揭示了超过哪些 A^2/σ^2 取值后，某一阶数 m 便不再是最优选择。图 10b 比较了优化后的互信息估计值与星座元素服从均匀分布时所产生的互信息。显然，对于所考虑的每个信噪比值，学习得到的概率整形都能带来更高的互信息。

学习得到的概率整形也对应于 [49]中推导出的最优输入分布的渐近取值。具体而言，研究表明，当 $\{-A, A\}$ 时，最优分布会收敛于 $A^2/\sigma^2 \to 0$上的伯努利分布；而当 $A^2/\sigma^2 \to \infty$ 时，概率质量函数在整个区间 $[-A, A]$上变为均匀分布。图11 展示了 $k = 16$时学习得到的概率整形，确实符合上述渐近行为。中间点 $A^2/\sigma^2 = 12[dB]$ 则展示了在两种极端情况之间进行插值的情形。



(a) $A^2/\sigma^2 = -10[dB]$。 (b) $A^2/\sigma^2 = 15[dB]$ (c) $A^2/\sigma^2 = 30[dB]$

图11：针对阶数为 $k = 16$ 且 A^2/σ^2 取若干不同值时学习得到的概率整形。



(a) 信息速率对比。 (b) 与均匀分布的对比。

图 12：QAM概率整形性能。图 (a) 展示了多个星座阶数下优化后互信息与解析上界和下界的对比，分别来自 [50]和 [51], ；图 (b) 展示了优化后互信息与均匀分布所产生的互信息之间的对比。

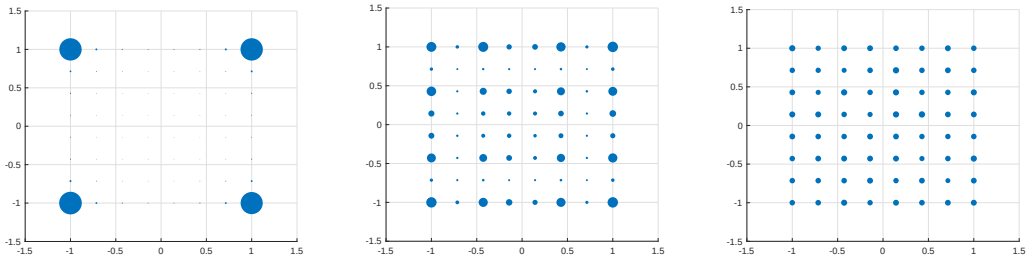
7.2 复加性高斯白噪声

针对复加性高斯白噪声信道的编码，我们考虑矩形QAM星座。此时，峰值约束变为盒约束，其中 Re(X) 和 Im(X) 受一维峰值功率约束限制。在盒约束下， [52] 表明，最优输入 X 的 Re(X) 和 Im(X) 相互独立，这是由于噪声分量相互独立。利用这一事实，我们得到容量界

$$2C_{\mathsf{LB}}^{\mathsf{real}} \leq C \leq 2C_{\mathsf{UB}}^{\mathsf{real}}, \tag{25}$$

其中 $C_{\mathsf{LB}}^{\mathsf{real}}$和 $C_{\mathsf{UB}}^{\mathsf{real}}$是第 7.1 节中的实值峰值功率加性高斯白噪声容量界 .

图12a 比较了我们估计的互信息值与上述界，并涵盖了多种正交幅度调制阶数。图12b 中，我们将互信息神经估计器的估计值与由均匀分布诱导的互信息进行比较，后者通过高斯-埃尔米特积分近似计算得出。显然，对于所有所考虑的 A^2/σ^2 值，学习到的分布都优于均匀分布。最后，图13 展示了 $k = 64$ 以及 A^2/σ^2的若干取值下学习到的概率整形。由于 Re(X) 和Im(X) 相互独立，我们寻找与实值加性高斯白噪声信道中观察到的特征类似的特征。确实，我们发现，当 A^2/σ^2 取值较低时，学习到的概率整形在星座边缘上呈均匀分布；而随着 A^2/σ^2 增大，概率整形逐渐转向在整个星座上均匀分布。对于13(b)所示的中间取值 A^2/σ^2，可以观察到非平凡的输入分布。



(a) $A^2/\sigma^2 = -10[dB]$。 (b) $A^2/\sigma^2 = 12[dB]$ (c) $A^2/\sigma^2 = 30[dB]$

图13：对于若干取值的 A^2/σ^2 和 $k = 64$，学习得到的正交幅度调制分布。标记大小表示分配的概率。

8 证明

8.1 定理的证明 1

证明表明：(i) Z_t 随（Z_{t-1}，U_{t-1}，W_{t-1}）变化；(ii) $P_{W_t|W^{t-1},\ U^t,\ Z^t} = P_{W_t|W_{t-1},\ U_t,\ Z^t}$；以及 (iii) $\rho(\pi_\phi) = \mathsf{I}_\phi(\mathbb{X} \to \mathbb{Y})$。首先，通过将 f 定义为 Z_{t-1} 与（U_t，W_t）的拼接，可得函数关系 $Z_t = f(Z_{t-1}, U_t, W_t)$。对于 (ii)，我们有

$$\begin{aligned} P_W(W_t = w|W^{t-1} = w^{t-1}, Z^{t-1} = z^{t-1}, U^{t-1} = u^{t-1}) &= \mathbb{P}(Y_0 = y|X_{-t}^0 = x_{-t}^0, Y_{-t}^0 = y_{-t}^0) \\ &= P_W(W_t = w|Z_{t-1} = z_{t-1}, U_{t-1} = u_{t-1}). \end{aligned}$$

为证明 (iii)，注意到

$$\begin{aligned} \rho(\pi_\phi) &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E}\left[\mathbb{E}\left[\log \frac{P_{Y_0|Y_{-t}^{-1}, X_{-t}^0}}{P_{Y_0|Y_{-t}^{-1}}} \middle| X_{-t}^0, Y_{-t}^{-1}\right]\right] \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathsf{I}_\phi(X_{-t}^0; Y_0|Y_{-t}^{-1}) \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathsf{I}_\phi(X^t; Y_t|Y^{t-1}) \\ &= \lim_{N\to\infty} \frac{1}{N} \mathsf{I}_\phi(X^N \to Y^N) \\ &= \mathsf{I}_\phi(\mathbb{X} \to \mathbb{Y}), \end{aligned}$$

其中第三个等号利用了联合分布的平稳性。证明至此结束。□

引理 1 的证明

回顾一下，我们关注的损失函数为:

$$\widehat{\mathsf{J}}_{\theta}^{\mathsf{MI}}(D_n,\phi) := \frac{1}{n}\sum_{t=1}^{n}\log\left(p^{\phi}(X_t)\right)\left(g_{\theta}(X_t,Y_t)-\widehat{\mathsf{I}}_{\mathsf{MI}}(D_n)\right). \tag{26}$$

首先考虑关于单个坐标的偏导数。不失一般性，假设 $\mathcal{X}=[1,\ldots,m]$ 并固定 $i\in[1,\ldots,m]$。对 ϕ_i 求导，得到

$$\partial_{\phi_i}\widehat{\mathsf{J}}_{\theta}^{\mathsf{MI}}(D_n,\phi) = \frac{1}{n}\sum_{t=1}^{n}\partial_{\phi_i}\log\left(p^{\phi}(X_t)\right)\left(g_{\theta}(X_t,Y_t)-\widehat{\mathsf{I}}_{\mathsf{MI}}(D_n)\right).$$

由于 $p^{\phi}(X_t)=\sigma^{X_t}{}_{\mathsf{sm}}(\phi)$，我们得到

$$\begin{aligned}\partial_{\phi_i}\log p^{\phi}(X_t) &= \partial_{\phi_i}\log e^{\phi_{X_t}} - \partial_{\phi_i}\log\left(\sum_{k=1}^{m}e^{\phi_k}\right)\\ &= \partial_{\phi_i}\phi_{X_t} - \frac{e^{\phi_i}}{\sum_{k=1}^{m}e^{\phi_k}}\\ &= {}_{\mathbb{I}}\{X_t=i\} - \sigma^{X_t}_{\mathsf{sm}}(\phi).\end{aligned}$$

因此，梯度为

$$\nabla_{\phi}\widehat{\mathsf{J}}_{\theta}^{\mathsf{MI}}(D_n,\phi) = \frac{1}{n}\sum_{t=1}^{n}\left(e_{X_t}-p^{\phi}\right)\left(g_{\theta}(X_t,Y_t)-\widehat{\mathsf{I}}_{\mathsf{MI}}(D_n)\right). \tag{27}$$

□

引理 2 的证明

为简化符号，令 $m:=|\mathcal{Q}_g|$ 并固定 m。注意，在 y 标记下相同的图的数量为 $(m)^{|\mathcal{Y}|m}$。文献 [18] 的作者声称，图池化方法可将算法考虑的 Q-图数量减少 $m!$ 倍。因此，

$$N_{\mathsf{GP}} = \frac{m^{m|\mathcal{Y}|}}{m!}. \tag{28}$$

根据 $m!$ 的斯特林上界近似，我们有

$$\begin{aligned}N_{\mathsf{GP}} &\geq \frac{m^{m|\mathcal{Y}|}}{\sqrt{e^2 m}\frac{m^m}{e^m}}\\ &= e^m m^{m(|\mathcal{Y}|-1)-0.5}e^{-1}\\ &\geq e^m m^{m-0.5}\end{aligned}$$

$$\geq e^{m((m-0.5)\log m)}$$
$$\geq e^{m\log m}.$$

9 总结与未来方向

本研究提出了一种针对具有离散输入字母表的通信信道优化估计互信息速率的方法。我们为输入概率质量函数提出了一种深度生成模型，并推导出一个替代优化目标，该目标易于针对概率质量函数模型的参数求导。我们通过对互信息优化问题进行马尔可夫决策过程表述，结合策略梯度方法和基于 DINE 的函数逼近，推导出了这一新目标。整体流程是对互信息进行迭代估计优化，其中估计步骤包括训练 DINE。据我们所知，这是首个在信道模型未知时，能够针对离散输入优化估计互信息的方法。

为展示我们方法的实用性，我们将其用于估计各种信道在前向通信方案和反馈通信方案下的容量。容量估计值与已知的理论解和/或界显著吻合，并且学习得到的输入概率质量函数与达到容量的输入分布一致。此外，我们展示了如何利用优化后的输入概率质量函数模型，通过 Q-图计算单叶有限状态信道反馈容量的下界和上界。最后，我们展示了该算法如何为 PP-AWGN 的脉冲幅度调制和正交幅度调制星座构造概率整形方案。

我们的工作支持对估计的互信息进行优化，并将信道视为可进行采样的黑盒。当假设可以控制其某些元素时，这种方法有益于数据驱动的时间序列任务。未来，我们计划将所提出的方案应用于多用户通信信道，并将我们的框架推广到多智能体强化学习。我们还将探索其在序列机器学习和序列控制中的应用。

参考文献

[1] James Massey。因果性、反馈与定向信息。载于《国际信息理论应用研讨会论文集》（ISITA-90），第 303–305 页。Citeseer，1990。[2] Gerhard Kramer。带反馈信道的定向信息，第 11 卷。Citeseer，1998。[3] Haim H Permuter、Young-Han Kim 和 Tsachy Weissman。论定向信息与赌博。载于《2008 年 IEEE 信息理论国际研讨会论文集》，第 1403–1407 页。IEEE，2008。[4] Arvind Rao、Alfred O Hero、David J States 和 James Douglas Engel。利用定向信息准则推断具有生物学意义的基因影响网络。载于《2006 年 IEEE 声学、语音与信号处理国际会议论文集》，第 2 卷，第 II–II 页。IEEE，2006。[5] Michael Wibral、Raul Vicente 和 Joseph T Lizier。神经科学中的定向信息度量。Springer，2014。[6] Yuxun Zhou 和 Costas J Spanos。因果性与次模性的结合：基于定向信息的子集选择。载于《神经信息处理系统进展》，第 2649–2657 页，2016。[7] Stas Tiomkin 和 Naftali Tishby。马尔可夫决策过程中的因果信息与价值的统一贝尔曼方程。arXiv 预印本 arXiv:1703.01585，2017。[8] Young-Han Kim。一类带反馈平稳信道的编码定理。《IEEE 信息论汇刊》，54(4):1488–1499，2008。[9] Young-Han Kim。一阶移动平均高斯信道的反馈容量。《IEEE 信息论汇刊》，52(7):3063–3079，2006。[10] Haim Permuter、Paul Cuff、Benjamin Van Roy 和 Tsachy Weissman。陷阱门信道的反馈容量。《IEEE 信息论汇刊》，54(7):3150–3165，2008。[11] Oron Sabag、Haim H Permuter 和 Navin Kashyap。具有无连续 1 输入约束的二元擦除信道的反馈容量。《IEEE 信息论汇刊》，62(1):8–22，2015。[12] Ori Peled、Oron Sabag 和 Haim H Permuter。（0, k）-RLL 输入约束 BEC 的反馈容量与编码。《IEEE 信息论汇刊》，65(7):4097–4114，2019。[13] Iddo Naiss 和 Haim H Permuter。扩展 Blahut–Arimoto 算法以最大化定向信息。《IEEE 信息论汇刊》，59(1):204–222，2012。

[14] Charalambos D Charalambous 和 Photios A Stavrou。《抽象空间上的定向信息：性质与变分等式》。《IEEE 信息论汇刊》，62(11):6019–6052，2016。[15] Ohad Elishco 和 Haim Permuter。《带反馈伊辛信道的容量与编码》。《IEEE 信息论汇刊》，60(9):5138–5149，2014。[16] Ziv Aharoni、Oron Sabag 和 Haim Henri Permuter。《大字母表伊辛信道反馈容量的强化学习评估与求解》。arXiv 预印本 arXiv:2008.07983，2020。[17] Andrea Grigorescu、Holger Boche、Rafael F Schaefer 和 H Vincent Poor。《带反馈有限状态信道的容量：算法与优化理论性质》。arXiv 预印本 arXiv:2201.11639，2022。[18] Oron Sabag、Bashar Huleihel 和 Haim H Permuter。《基于图的编码器及其在带反馈有限状态信道中的性能》。《IEEE 通信汇刊》，68(4):2106–2117，2020。[19] Bashar Huleihel、Oron Sabag、Haim H Permuter、Navin Kashyap 和 Shlomo Shamai。《有限状态信道容量的可计算上界》。《IEEE 信息论汇刊》，2021。[20] Nunzio A Letizia 和 Andrea M Tonello。《面向通信的容量驱动自编码器》。《IEEE 通信学会开放期刊》，2021。[21] Farhad Mirkarimi、Stefano Rini 和 Nariman Farsad。《神经网络容量估计器：其可靠性如何？》arXiv 预印本 arXiv:2111.07401，2021。[22] Dor Tsur、Ziv Aharoni、Ziv Goldfeld 和 Haim Permuter。《连续空间上定向信息的神经网络估计与优化》。arXiv 预印本 arXiv:2203.14743，2022。[23] Richard S Sutton、David A McAllester、Satinder P Singh、Yishay Mansour 等。《带函数逼近的强化学习策略梯度方法》。载于 NIPs，第99卷，第1057–1063页。Citeseer，1999。[24] Oron Sabag、Haim H Permuter 和 Henry D Pfister。《单叶有限状态信道反馈容量的单字母上界》。《IEEE 信息论汇刊》，63(3):1392–1409，2016。[25] John G Proakis、Masoud Salehi、Ning Zhou 和 Xiaofeng Li。《通信系统工程》，第2卷。Prentice Hall，新泽西州，1994。[26] Thomas M Cover 和 A Joy Thomas。《信息论基础》。Wiley，纽约，第2版，2006。[27] Roland L Dobrushin。《信息论中香农主定理的一般表述》。《美国数学学会译丛》，33:323–438，1963。

[28] Monroe D Donsker 和 SR Srinivasa Varadhan。《大时间下某些马尔可夫过程期望的渐近评估，IV》。《纯粹与应用数学通讯》，36(2):183–212，1983。[29] Liang Jin、Madan M Gupta 和 Peter N Nikiforuk。《利用动态循环神经网络实现通用逼近：离散时间版本》。载于 ICNN'95——国际神经网络会议，第1卷，第403–408页。IEEE，1995。[30] Dimitri P Bertsekas 等。《动态规划与最优控制：第1卷》。Athena scientific，贝尔蒙特，2000。[31] Sekhar Tatikonda 和 Sanjoy Mitter。《带反馈信道的容量》。《IEEE 信息论汇刊》，55(1):323–349，2008。[32] Rolando Cavazos-Cadena 和 Linn I Sennott。《比较近期关于平均最优平稳策略存在性的假设》。《运筹学快报》，11(1):33–37，1992。[33] Aristotle Arapostathis、Vivek S Borkar、Emmanuel Fern´andez-Gaucherand、Mrinal K Ghosh 和 Steven I Marcus。《具有平均成本准则的离散时间受控马尔可夫过程：综述》。《SIAM 控制与优化杂志》，31(2):282–344，1993。[34] Sheldon M Ross。《随机动态规划导论》。学术出版社，2014。[35]Li Xia、Xianping Guo 和 Xi-Ren Cao。《可数状态平均马尔可夫决策过程中最优平稳策略的存在性》。arXiv 预印本 arXiv:2007.01602，2020。[36] Eli Shemuel、Oron Sabag 和 Haim Permuter。《噪声输出即状态（NOST）信道的反馈容量》，2021。[37] Timothy P Lillicrap、Jonathan J Hunt、Alexander Pritzel、Nicolas Heess、Tom Erez、Yu-val Tassa、David Silver 和 Daan Wierstra。《利用深度强化学习进行连续控制》。arXiv 预印本 arXiv:1509.02971，2015。[38] Mohamed Ishmael Belghazi、Aristide Baratin、Sai Rajeshwar、Sherjil Ozair、Yoshua Ben-gio、Aaron Courville 和 Devon Hjelm。《互信息神经估计》。载于国际机器学习会议，第531–540页。PMLR，2018。[39] Sreejith Sreekumar、Zhengxin Zhang 和 Ziv Goldfeld。《神经网络估计 f-散度的非渐近性能保证》。载于人工智能与统计国际会议，第3322–3330页。PMLR，2021。[40] Sreejith Sreekumar 和 Ziv Goldfeld。《统计散度的神经网络估计》。《机器学习研究杂志》，23(126):1–75，2022。

[41] Sepp Hochreiter 和 J¨urgen Schmidhuber。长短期记忆。《神经计算》，9(8):1735–1780，1997。[42] Haim Henry Permuter、Tsachy Weissman 和 Andrea J Goldsmith。具有时不变确定性反馈的有限状态信道。《IEEE 信息论汇刊》，55(2):644–662，2009。[43] Mordechai Mushkin 和 Israel Bar-David。Gilbert-Elliott信道的容量与编码。《IEEE 信息论汇刊》，35(6):1277–1290，1989。[44] Mohammad Rezaeian。使用统计方法计算Gilbert-Elliott信道的容量。载于《2005年澳大利亚通信理论研讨会》，第56–61页。IEEE，2005。[45] Bashar Huleihel、Oron Sabag 和 Haim H Permuter。具有延迟反馈的陷阱门信道容量。载于《2022年IEEE国际信息论研讨会（ISIT）》，第492–497页。IEEE，2022。[46] Kingo Kobayashi。陷阱门信道的容量问题。载于《信息传输与组合学通论》，第1084–1087页。Springer，2006。[47]Haim Henri Permuter、Himanshu Asnani 和 Tsachy Weissman。有反馈和无反馈时邮政信道的容量。《IEEE 信息论汇刊》，60(10):6041–6057，2014。[48] Anil K Jain 和 Richard C Dubes。数据聚类算法。Prentice-Hall公司，1988。[49] Joel G Smith。幅度与方差受限标量高斯信道的信息容量。《信息与控制》，18(3):203–219，1971。[50] Lawrence H Ozarow 和 Aaron D Wyner。关于具有有限个输入电平的高斯信道容量。《IEEE 信息论汇刊》，36(6):1426–1428，1990。[51] Andrew Thangaraj、Gerhard Kramer 和 Georg B¨ocherer。离散时间、幅度受限加性高斯白噪声信道的容量界。《IEEE 信息论汇刊》，63(7):4172–4182，2017。[52] Shiro Ikeda、Kazunori Hayashi 和 Toshiyuki Tanaka。峰值功率约束下的容量与调制。arXiv预印本 arXiv:1005.3889，2010。