



中国科技期刊卓越行动计划项目入选期刊

控制与决策

C O N T R O L A N D D E C I S I O N



基于物理信息神经网络的随机系统演化分析

曹瑞, 裔扬, 刘燕斌

引用本文:

曹瑞, 裔扬, 刘燕斌. 基于物理信息神经网络的随机系统演化分析[J]. *控制与决策*, 2024, 39(9): 3013–3022.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0747>

您可能感兴趣的其他文章

Articles you may be interested in

[基于数据驱动的非线性网络系统自适应迭代学习控制](#)

Data driven adaptive learning control of nonlinear network system

控制与决策. 2021, 36(6): 1523–1528 <https://doi.org/10.13195/j.kzyjc.2019.1182>

[基于神经网络的电力系统暂态稳定分布式自适应控制](#)

Neural network-based distributed adaptive control for power system transient stability

控制与决策. 2021, 36(6): 1407–1414 <https://doi.org/10.13195/j.kzyjc.2019.1168>

[基于强化学习的倒立摆分数阶梯度下降RBF控制](#)

Reinforcement learning based fractional gradient descent RBF neural network control of inverted pendulum

控制与决策. 2021, 36(1): 125–134 <https://doi.org/10.13195/j.kzyjc.2019.0816>

[工业信息物理系统安全风险动态表现分析量化评估模型](#)

Quantitative evaluation model for dynamic performance analysis of security risk in industrial cyber physics systems

控制与决策. 2021, 36(8): 1939–1946 <https://doi.org/10.13195/j.kzyjc.2019.1479>

[具有未建模动态和输出约束的耦合系统的分散自适应控制](#)

Decentralized adaptive control for interconnected systems with unmodeled dynamics and output constraints

控制与决策. 2018, 33(12): 2113–2121 <https://doi.org/10.13195/j.kzyjc.2017.0933>

基于物理信息神经网络的随机系统演化分析

曹瑞^{1†}, 裔扬¹, 刘燕斌²

(1. 扬州大学 信息工程学院, 江苏 扬州 225127; 2. 南京航空航天大学 航天学院, 南京 211106)

摘要: 提出一个概率深度学习框架, 通过基于物理信息的神经网络对非线性微分系统中的不确定性进行量化和演化分析. 算法使用潜在变量模型(即噪声/随机变量模型)构建系统状态的概率表示, 并将偏微分方程描述的物理定律引入训练过程中, 建立约束关系, 为基于小样本数据集的训练提供一种有效的机制. 所提出的概率物理学信息神经网络算法能够提供一种灵活的框架, 表征由输入中的随机性或观测中的噪声引起的物理系统输出的不确定性, 避免重复采样昂贵的实验或数值模拟器的需要. 最后, 通过包含随机变量的非线性系统示例对所提出算法进行验证, 以表明算法的有效性.

关键词: 物理信息; 神经网络; 随机系统; 小样本数据; 概率神经网络; 演化分析

中图分类号: TP273 文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0747

引用格式: 曹瑞, 裔扬, 刘燕斌. 基于物理信息神经网络的随机系统演化分析 [J]. 控制与决策, 2024, 39(9): 3013-3022.

Evolutionary analysis of stochastic systems based on physical information neural networks

CAO Rui^{1†}, YI Yang¹, LIU Yan-bin²

(1. College of Information Engineering, Yangzhou University, Yangzhou 225127, China; 2. College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: This study presents a probabilistic deep learning framework that analyzes the uncertainty in nonlinear differential systems using a physics-informed neural network. The framework utilizes a latent variable model (i.e. noise/random variable model), to create a probability-based model of the system state. Physical laws that are described using partial differential equations are incorporated into the training process to establish constraint relationships, which provides an effective way to train using small samples. The proposed probabilistic physics information neural network (PPINN) algorithm provides a flexible framework, representing the uncertainty of the physical system output caused by the randomness in the input or the noise in the observation, which helps avoid the high costs associated with repeated sampling or expensive experiments. Finally, the effectiveness of the proposed algorithm is demonstrated through some nonlinear system examples containing random variables.

Keywords: physical information; neural network; stochastic system; small sample data; probabilistic neural network; evolution analysis

0 引言

机器学习和数据分析的最新进展在不同的科学学科中产生了变革性的结果, 并且由于传感器、数据存储和计算资源的可用性, 使得数据驱动方法在许多学科中占据了重要地位^[1]. 对于数据驱动学习而言, 当使用大量数据进行训练时, 所有解决方案都能获得最先进的性能. 然而, 在分析和预测复杂现象的实验

中, 获得大量高质量实验数据的成本高得令人望而却步. 这种情况下, 对于系统的复杂性而言, 当数据相对稀缺时纯数据驱动方法会带来困难. 这是因为绝大多数机器学习技术缺乏鲁棒性, 难以提供收敛保证, 或无法量化与预测相关的误差/不确定性. 因此, 在数据有限的领域中, 具有稳健和高效采样方式学习的能力是必要的, 这就要求利用特定领域的知识对数据学

收稿日期: 2023-05-30; 录用日期: 2023-08-02.

基金项目: 江苏省自然科学基金青年项目(BK20230560); 国家自然科学基金项目(62303400, 52272369).

责任编辑: 柴利.

[†]通讯作者. E-mail: stdio@yzu.edu.cn.

*本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

习类算法进行约束,使其返回能够满足某些物理原理(如质量守恒、动量守恒等)的预测。

纯数据驱动学习算法的缺点经常引起应用数学家和工程师对此类方法的怀疑。在最近的工作中,Raissi等^[2-3]重新审视Lagaris等^[4]提出的使用微分方程约束惩罚深度神经网络损失函数的想法,踏上了应用数学与机器学习相结合的道路。这一系列工作已从经验上证明了这种基于物理信息的神经网络(physics-informed neural networks, PINN)方法如何在小数据范围内规范学习,以在不完整模型或不完整数据中预测复杂系统的动力学。尽管在规范问题上取得了一系列令人印象深刻的结果,但Raissi等^[2-3]也指出了PINN算法缺乏关于收敛和后验误差估计的支持理论。此外,PINN算法通过数据和先验信息对确定性物理系统进行描述,尚未涉及随机系统,而在实际系统中,建模误差以及环境干扰对动力学带来的不确定影响也需要进行分析和研究,这自然需要对PINN算法进行扩展以对不确定性进行量化。

当前不确定性量化技术主要分为两个方面。一方面,应用数学和科学计算的研究人员主要依赖于源于第一物理原理的数学模型,旨在阐明系统行为的关键机制,同时表征由此产生的预测在解释不确定性方面的稳健性^[5]。尽管此类方法在过去20年中取得了很大进展,但在计算随机系统的最流行方法中,如多项式混沌展开^[6]、稀疏网格求积^[7]、多级/多保真蒙特卡罗采样^[8]、正交分解^[9]和高斯过程回归模型^[10],现实应用中遇到非高斯似然和高维后验分布时还是会面临着严重的限制。另一方面,机器学习研究的爆炸式增长为学习和操纵复杂的高维概率分布提供了新的有效方法。相关推理工具,如变分自动编码器^[11]和生成对抗性网络^[12],在灵活网络模块(如前馈/卷积/递归神经网络^[13])的基础上制定,已经出现了可扩展的解决方案,处理假设先验信息不多的问题。这项工作与已有相关研究的创新之处在于,提出了一类概率物理学信息神经网络,不仅处理了非高斯分布对理论不确定性量化的限制,还通过偏微分方程表示的先验信息,降低机器学习对训练数据量的要求,能够直接从有噪声的输入/输出数据中训练此类模型,对含有随机变量的不确定系统进行演化分析。具体而言,本文为了达到上述目标,在算法中主要构建模块不能再是清晰的确定性数字和支配定律,而是必须使用概率模型。

本文首先对Raissi等^[2-3]提出的物理信息神经网络进行简要概述;然后对所提出的概率物理学信息神经网络(probabilistic physics information neural

network, PPINN)算法进行描述,并介绍所提出的概率公式和变分推理框架;最后,通过两个示例对所提出算法的有效性进行验证。

1 PINN算法概述

Raissi等^[2-3,14]的工作已经证明了经典守恒定律和数值离散化方案如何被用作结构化先验信息,以提高现代机器学习算法的鲁棒性和效率,并引入了一类新的基于物理的数据驱动学习方法来发现模型。在上述工作中,考虑构建深度神经网络,以返回受以下参数化偏微分方程(partial differential equations, PDE)约束的预测信息:

$$\mathbf{u}_t + \mathcal{N}_x \mathbf{u} = 0. \quad (1)$$

其中: $\mathbf{u}(\mathbf{x}, t)$ 代表系统输出,PINN利用 θ 参数化的深度神经网络对其进行描述,即 $\mathbf{u}(\mathbf{x}, t) = f_\theta(\mathbf{x}, t)$, \mathbf{x} 为空间坐标向量, t 为时间坐标, \mathcal{N}_x 为非线性微分算子。由于神经网络是可微的,可以通过式(1)建立约束,其对应于PDE残差,即 $r_\theta(\mathbf{x}, t) := \partial f_\theta(\mathbf{x}, t)/\partial t + \mathcal{N}_x f_\theta(\mathbf{x}, t)$ 。由此产生的训练过程允许使用系统的一些分散观测值 $\{(\mathbf{x}_i, t_i), u_i\} (i = 1, 2, \dots, N_u)$ 以及大量的配点 $\{(\mathbf{x}_i, t_i), r_i\} (i = 1, 2, \dots, N_r)$ 恢复网络参数 θ ,旨在惩罚 N_r 个配置节点的有限集合处的PDE残差。由式(1)可知,PDE残差数据为零,即 $r_i = 0$ 。这种简单但非常有效的正则化过程,允许在模型的似然函数中引入PDE残差作为软惩罚约束,并使用标准随机梯度下降法有效地解决由此产生的优化问题,而不需要任何复杂的约束优化技术^[2-3],简单地通过最小化如下复合损失函数即可获得满足要求的网络预测模型 $f_\theta(\mathbf{x}, t)$:

$$\mathcal{L}(\theta) = \frac{1}{N_u} \sum_{i=1}^{N_u} \|f_\theta(\mathbf{x}_i, t_i) - u_i\|^2 + \frac{1}{N_r} \sum_{i=1}^{N_r} \|r_\theta(\mathbf{x}_i, t_i) - r_i\|^2, \quad (2)$$

其中 $\partial \mathcal{L}/\partial \theta$ 可以使用自动微分方法^[15]进行获取。

由于得到的网络预测模型能够满足PDE约束所施加的任何物理特性,这种方法展示了如何在机器学习算法中近似编码物理系统中的特定约束,从而引入一种新的形式对小样本数据集进行学习。

2 概率物理学信息神经网络算法

为了处理不确定性问题,提出概率物理学信息神经网络算法用于描述随机系统,并使用潜在变量模型(即噪声/随机变量)来构建系统状态的概率表示。PPINN算法提出了一个概率公式,用于通过基于物理的神经网络分析不确定性演化,构建形式如下:

$$\begin{aligned} p(\mathbf{u}|\mathbf{x}, t, \mathbf{z}), \mathbf{z} \sim p(\mathbf{z}), \\ \text{s.t. } \mathbf{u}_t + \mathcal{N}_x \mathbf{u} = 0. \end{aligned} \quad (3)$$

式(3)封装了一系列确定性和随机性问题,其中 $\mathbf{u}(\mathbf{x}, t, \mathbf{z})$ 是一个包含潜在变量的随机网络,其主要取决于空间变量 \mathbf{x} 和时间变量 t ,潜在随机变量 \mathbf{z} 是满足先验分布 $p(\mathbf{z})$ 的集合,作用是总结在可观测输出 \mathbf{u} 中引入随机潜在在高维因素的影响. 本文假设用于预测 $\mathbf{u}(\mathbf{x}, t, \mathbf{z})$ 的数据通常仅包含PDE的初始/边界条件,或者随机系统所感兴趣的区域内收集的稀疏观测. 所提出的PPINN算法与现有的不确定性量化流行方法相比,不需要先验地指定随机因素;此外,随机变量对观测输出引入的影响是通过潜在变量 \mathbf{z} 提取的.

从数据中学习随机系统模型的能力是概率科学计算和物理系统不确定性量化的基石. 受物理信息约束的条件概率 $p(\mathbf{u}|\mathbf{x}, t, \mathbf{z})$ 引入了一种正则化机制,该机制将解空间限制在可容许范围内(例如,在流体力学问题中会丢弃违反质量守恒原理的非现实解),从而能够在小样本数据状态下训练概率深度网络模型. 在这里,“小样本数据”一词应该相对于底层系统的复杂性来考虑. 另外需要说明的是,这里使用“小样本数据”的概念是用来强调所提出的方法解决这一具有挑战性的问题的能力,该方法能够使用相对较小的测量集来训练概率网络模型,绕过了大量采样模拟器以表征系统响应的需要.

2.1 联合分布匹配的对抗性推理

根据文献[16]的研究结果,通过最小化反向Kullback-Leibler(KL)散度 $\text{KL}[p_\theta(\mathbf{x}, t, \mathbf{u})||q(\mathbf{x}, t, \mathbf{u})]$ 能够用来匹配网络模型生成数据的概率分布 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 和观测数据的概率分布 $q(\mathbf{x}, t, \mathbf{u})$,以使网络模型生成的概率分布足够近似观测的实际数据,这是训练式(3)中网络模型比较有前途的一种方法. 使用反向KL散度的计算表达式如下:

$$\begin{aligned} \text{KL}[p_\theta(\mathbf{x}, t, \mathbf{u})||q(\mathbf{x}, t, \mathbf{u})] = \\ -h(p_\theta(\mathbf{x}, t, \mathbf{u})) - E_{p_\theta(\mathbf{x}, t, \mathbf{u})}[\log(q(\mathbf{x}, t, \mathbf{u}))]. \end{aligned} \quad (4)$$

等式右侧第1项 $h(p_\theta(\mathbf{x}, t, \mathbf{u}))$ 表示生成模型的熵,第2项可以分解为

$$\begin{aligned} E_{p_\theta(\mathbf{x}, t, \mathbf{u})}[\log(q(\mathbf{x}, t, \mathbf{u}))] = \\ \int \int \int_{\mathcal{S}_{p_\theta} \cap \mathcal{S}_q} \log(q(\mathbf{x}, t, \mathbf{u})) p_\theta(\mathbf{x}, t, \mathbf{u}) d\mathbf{x} dt d\mathbf{u} + \\ \int \int \int_{\mathcal{S}_{p_\theta} \cap \mathcal{S}_q^c} \log(q(\mathbf{x}, t, \mathbf{u})) p_\theta(\mathbf{x}, t, \mathbf{u}) d\mathbf{x} dt d\mathbf{u}, \end{aligned} \quad (5)$$

其中 \mathcal{S}_{p_θ} 和 \mathcal{S}_q 分别为分布 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 和 $q(\mathbf{x}, t, \mathbf{u})$ 的支持区域, \mathcal{S}_q^c 为 \mathcal{S}_q 的补码. 在式(4)的反向Kullback-

Leibler散度计算中引入一种机制,以试图平衡两个竞争目标的影响. 即最大化熵 $h(p_\theta(\mathbf{x}, t, \mathbf{u}))$ 是鼓励 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 在其支持集上广泛分布,而式(5)中的第2个积分项会在 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 和 $q(\mathbf{x}, t, \mathbf{u})$ 的支持集不重叠时引入强烈的惩罚,从而鼓励 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 的支持集仅扩展到 $\mathcal{S}_{p_\theta} \cap \mathcal{S}_q^c = \emptyset$ 的点,这意味着 $\mathcal{S}_{p_\theta} \subseteq \mathcal{S}_q$. 而当 $\mathcal{S}_{p_\theta} \subset \mathcal{S}_q$ 时,又会使数据集的多样性不够,导致“模式崩溃”的病理性问题[17]. 这一问题也是生成对抗网络的训练中普遍存在的问题[12],若直接最小化式(4)中的反向KL散度,则会出现上述的“模式崩溃”病理性问题,因为仅依靠式(4)无法控制两个项的相对重要性.

需要说明的是,式(5)所列的分解在这里仅用于激励对 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 支持集的约束,并未实际用于本文算法中. 将其列出是为了提供关于“模式崩溃”病理的理解. 为了减轻“模式崩溃”的程度,可以通过在式(4)中引入权重值 $\lambda \geq 1$,即 $-\lambda h(p_\theta(\mathbf{x}, t, \mathbf{u})) - E_{p_\theta(\mathbf{x}, t, \mathbf{u})}[\log(q(\mathbf{x}, t, \mathbf{u}))]$ 进行最小化. 这是由于 $h(p_\theta(\mathbf{x}, t, \mathbf{u}))$ 引入的熵正则化能够为减轻模式崩溃的影响提供一种有效的机制,从而增强学习 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 对抗性推理的鲁棒性.

对式(4)进行最小化存在两个基本困难. 首先,评估 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 和 $q(\mathbf{x}, t, \mathbf{u})$ 这两个分布会涉及到高维的不可计算积分,故仅能得到这两个分布的样本,无法得到其显式解析形式. 其次,熵项 $h(p_\theta(\mathbf{x}, t, \mathbf{u}))$ 的微分不可计算,因为 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 不是先验已知的. 在下一节中,将重新审视文献[16]提出的无监督公式,并导出一个可行的推理过程,从 $\mathbf{u}(\mathbf{x}, t)$ 的分散观测数据 $(\{\mathbf{x}_i, t_i, \mathbf{u}_i\}, i = 1, 2, \dots, N_u)$ 对 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 进行学习.

2.1.1 密度比期望估计

根据定义,式(4)中反向KL散度涉及到计算对数密度比的期望,即

$$\begin{aligned} \text{KL}[p_\theta(\mathbf{x}, t, \mathbf{u})||q(\mathbf{x}, t, \mathbf{u})] := \\ E_{p_\theta(\mathbf{x}, t, \mathbf{u})} \left[\log \left(\frac{p_\theta(\mathbf{x}, t, \mathbf{u})}{q(\mathbf{x}, t, \mathbf{u})} \right) \right]. \end{aligned} \quad (6)$$

一般而言,给定来自两个分布的样本,可以通过构建一个二元分类器来近似它们的密度比,进而该分类器能够区分来自两个分布的样本. 为此,假设从 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 中抽取 N 个数据点,并分配一个标签 $y = +1$. 类似地,假设从 $q(\mathbf{x}, t, \mathbf{u})$ 中也抽取 N 个样本,并分配标签 $y = -1$. 可以将这些概率写成条件形式,即

$$\begin{aligned} p_\theta(\mathbf{x}, t, \mathbf{u}) &= \rho(\mathbf{x}, t, \mathbf{u}|y = +1), \\ q(\mathbf{x}, t, \mathbf{u}) &= \rho(\mathbf{x}, t, \mathbf{u}|y = -1), \end{aligned} \quad (7)$$

其中 $\rho(\mathbf{x}, t, \mathbf{u}|y = +1)$ 和 $\rho(\mathbf{x}, t, \mathbf{u}|y = -1)$ 为二元分类器 $T(\mathbf{x}, t, \mathbf{u})$ 预测的类概率. 使用贝叶斯定则, 可以容易地计算出 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 和 $q(\mathbf{x}, t, \mathbf{u})$ 的密度比为

$$\begin{aligned} \frac{p_\theta(\mathbf{x}, t, \mathbf{u})}{q(\mathbf{x}, t, \mathbf{u})} &= \frac{\rho(\mathbf{x}, t, \mathbf{u}|y = +1)}{\rho(\mathbf{x}, t, \mathbf{u}|y = -1)} = \\ &= \frac{\rho(y = +1|\mathbf{x}, t, \mathbf{u})\rho(\mathbf{x}, t, \mathbf{u})}{\rho(y = -1|\mathbf{x}, t, \mathbf{u})\rho(\mathbf{x}, t, \mathbf{u})} = \\ &= \frac{\rho(y = +1)}{\rho(y = -1)} = \\ &= \frac{\rho(y = +1|\mathbf{x}, t, \mathbf{u})}{\rho(y = -1|\mathbf{x}, t, \mathbf{u})} = \frac{\rho(y = +1|\mathbf{x}, t, \mathbf{u})}{1 - \rho(y = +1|\mathbf{x}, t, \mathbf{u})} = \\ &= \frac{T(\mathbf{x}, t, \mathbf{u})}{1 - T(\mathbf{x}, t, \mathbf{u})}. \end{aligned} \quad (8)$$

当 $\rho(y = +1) = \rho(y = -1) = 1/2$ 且网络模型的数据量与源自真实经验分布的数据量平衡时, 这两项会被抵消掉. 在模型训练期间, 可以随意选择生成数据的数量, 并将其设置为训练数据的大小. 而当网络模型生成数据与真实数据的数量不平衡时, $\rho(y = +1)/\rho(y = -1)$ 也只会式(8)中引入一个标量因子. 该过程提示了可以利用深度神经网络分类器的力量, 直接从数据中获得式(4)中的反向KL散度的准确估计, 而无需预先假设任何特定网络模型的概率分布 $p_\theta(\mathbf{x}, t, \mathbf{u})$.

2.1.2 熵正则化边界

按照文献[16]的推导方法, 构建一个计算熵 $h(p_\theta(\mathbf{x}, t, \mathbf{u}))$ 下界的表达式, 并基于此考虑包含随机变量 $(\mathbf{x}, t, \mathbf{u}, \mathbf{z})$ 的联合概率分布

$$\begin{aligned} p_\theta(\mathbf{x}, t, \mathbf{u}, \mathbf{z}) &= p_\theta(\mathbf{x}, t, \mathbf{u}|\mathbf{z})p(\mathbf{z}) = \\ &= p_\theta(\mathbf{u}|\mathbf{x}, t, \mathbf{z})p(\mathbf{x}, t, \mathbf{z}). \end{aligned} \quad (9)$$

其中: $p_\theta(\mathbf{u}|\mathbf{x}, t, \mathbf{z}) = \delta(\mathbf{u} - f_\theta(\mathbf{x}, t, \mathbf{z}))$, $\delta(\cdot)$ 为狄拉克 δ 函数. 随机变量 $(\mathbf{x}, t, \mathbf{u})$ 与 \mathbf{z} 之间的交互信息满足如下信息论恒等式:

$$\begin{aligned} I(\mathbf{x}, t, \mathbf{u}; \mathbf{z}) &= h(\mathbf{x}, t, \mathbf{u}) - h(\mathbf{x}, t, \mathbf{u}|\mathbf{z}) = \\ &= h(\mathbf{z}) - h(\mathbf{z}|\mathbf{x}, t, \mathbf{u}). \end{aligned} \quad (10)$$

其中: $h(\mathbf{x}, t, \mathbf{u})$ 、 $h(\mathbf{z})$ 为边缘熵, $h(\mathbf{x}, t, \mathbf{u}|\mathbf{z})$ 、 $h(\mathbf{z}|\mathbf{x}, t, \mathbf{u})$ 为条件熵^[18]. 本文设置 \mathbf{x} 和 t 是确定性变量, 独立于 \mathbf{z} , $p_\theta(\mathbf{u}|\mathbf{x}, t, \mathbf{z})$ 是由确定性函数 $f_\theta(\mathbf{x}, t, \mathbf{z})$ 生成的样本. 基于此可以得到 $h(\mathbf{x}, t, \mathbf{u}|\mathbf{z}) = 0$, 进而有

$$h(\mathbf{x}, t, \mathbf{u}) = h(\mathbf{z}) - h(\mathbf{z}|\mathbf{x}, t, \mathbf{u}), \quad (11)$$

其中 $h(\mathbf{z}) := -\int \log p(\mathbf{z})p(\mathbf{z})d\mathbf{z}$ 不依赖于网络模型参数 θ .

下面考虑由 ϕ 参数化的概率分布 $q_\phi(\mathbf{z}|\mathbf{x}, t, \mathbf{u})$, 其被用作潜在变量难以处理的后验概率 $p_\theta(\mathbf{z}|\mathbf{x}, t, \mathbf{u})$ 的变分近似, $h(\mathbf{z}|\mathbf{x}, t, \mathbf{u})$ 可以计算为

$$\begin{aligned} h(\mathbf{z}|\mathbf{x}, t, \mathbf{u}) &= \\ &= -E_{p_\theta(\mathbf{x}, t, \mathbf{u}, \mathbf{z})}[\log(p_\theta(\mathbf{z}|\mathbf{x}, t, \mathbf{u}))] = \\ &= -E_{p_\theta(\mathbf{x}, t, \mathbf{u}, \mathbf{z})}[\log(q_\phi(\mathbf{z}|\mathbf{x}, t, \mathbf{u}))] - \\ &= E_{p_\theta(\mathbf{x}, t, \mathbf{u})}[\text{KL}[p_\theta(\mathbf{z}|\mathbf{x}, t, \mathbf{u})||q(\mathbf{z}|\mathbf{x}, t, \mathbf{u})]] \leq \\ &= -E_{p_\theta(\mathbf{x}, t, \mathbf{u}, \mathbf{z})}[\log(q_\phi(\mathbf{z}|\mathbf{x}, t, \mathbf{u}))], \end{aligned} \quad (12)$$

其中 \mathbf{z} 被视为一组潜在变量. 若 $q_\phi(\mathbf{z}|\mathbf{x}, t, \mathbf{u})$ 作为与网络模型的概率分布 $p_\theta(\mathbf{z}|\mathbf{x}, t, \mathbf{u})$ 相关的辅助推理模型被引入, 其中 $\mathbf{u} = f_\theta(\mathbf{x}, t, \mathbf{z})$, $\mathbf{z} \sim p(\mathbf{z})$, 则可以使用式(11)和(12)限制方程(4)中的熵项, 即

$$\begin{aligned} h(p_\theta(\mathbf{x}, t, \mathbf{u})) &\geq \\ &= h(\mathbf{z}) + E_{p_\theta(\mathbf{x}, t, \mathbf{u})}[\log(q_\phi(\mathbf{z}|\mathbf{x}, t, \mathbf{u}))]. \end{aligned} \quad (13)$$

注意到, 推理模型 $q_\phi(\mathbf{z}|\mathbf{x}, t, \mathbf{u})$ (也称为编码器) 在潜在变量 \mathbf{z} 的真后验近似中扮演了一种变分逼近的角色, 在推导熵 $h(p_\theta(\mathbf{x}, t, \mathbf{u}))$ 下界的信息论论证中自然地出现.

2.2 对抗性训练的目标函数设置

利用第2.1.1节描述的密度比期望估计过程和第2.1.2节推导出的熵界, 可以得到如下损失函数:

$$\begin{aligned} \mathcal{L}_D(\psi) &= E_{q(\mathbf{x}, t)p(\mathbf{z})}[\log \sigma(T_\psi(\mathbf{x}, t, f_\theta(\mathbf{x}, t, \mathbf{z})))] + \\ &= E_{q(\mathbf{x}, t, \mathbf{u})p(\mathbf{z})}[\log(1 - \sigma(T_\psi(\mathbf{x}, t, \mathbf{u})))], \end{aligned} \quad (14)$$

$$\begin{aligned} \mathcal{L}_G(\theta, \phi) &= E_{q(\mathbf{x}, t)p(\mathbf{z})}[T_\psi(\mathbf{x}, t, f_\theta(\mathbf{x}, t, \mathbf{z})) + \\ &= (1 - \lambda) \log(q_\phi(\mathbf{z}|\mathbf{x}, t, f_\theta(\mathbf{x}, t, \mathbf{z})))]], \end{aligned} \quad (15)$$

进而通过熵正则化来最小化反向KL散度^[12,16]. 其中 $\sigma(x) = 1/(1 + e^{-x})$ 为逻辑符号函数, 这两种损失函数涉及数据点期望值.

式(14)和(15)已通过下标表示期望是以哪些分布进行采样的, 例如 $q(\mathbf{x}, t)p(\mathbf{z})$ 表示从经验公式的数据分布 $q(\mathbf{x}, t)$ 中采样 (\mathbf{x}, t) , 而 \mathbf{z} 是从先验 $p(\mathbf{z})$ 中采样. 式(14)本质上对应于二元分类中的交叉熵损失, 旨在改善分类器 $T_\psi(\mathbf{x}, t, \mathbf{u})$ 区分由网络模型 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 产生的“伪”样本 $(\mathbf{x}, t, f_\theta(\mathbf{x}, t, \mathbf{z}))$ 与源于观测数据分布 $q(\mathbf{x}, t, \mathbf{u})$ 的“真”样本 $(\mathbf{x}, t, \mathbf{u})$ 的能力. 式(15)的目标是改善生成器 $f_\theta(\mathbf{x}, t, \mathbf{z})$, 以生成越来越逼真的样本, 从而“欺骗”鉴别器 $T_\psi(\mathbf{x}, t, \mathbf{u})$. 具体而言, 式(15)中第1项通过使用式(8)中概述的密度比, 直接从样本中计算反向KL散度, 以估算 $p_\theta(\mathbf{x}, t, \mathbf{u})$ 和 $q(\mathbf{x}, t, \mathbf{u})$ 的联合分布匹配度. 式(15)第2项编码器 $q_\phi(\mathbf{z}|\mathbf{x}, t, f_\theta(\mathbf{x}, t, \mathbf{z}))$ 给出的分布不仅作为一个熵正则化机制, 还提供了对潜在变量 \mathbf{z} 真实后验的变分逼近.

式(15)试图使用以下过程将先验分布 $p(\mathbf{z})$ 生成的潜在变量和编码器预测的潜在变量与输入 $(\mathbf{x}, t, f_\theta(\mathbf{x}, t, \mathbf{z}))$ 进行匹配. 首先, 从先验分布 $p(\mathbf{z})$ 获取潜在变量, 通过生成器 $f_\theta(\mathbf{x}, t, \mathbf{z})$ 产生输出; 然后, 将输出传递给编码器 $q_\phi(\mathbf{z}|\mathbf{x}, t, f_\theta(\mathbf{x}, t, \mathbf{z}))$ 以预测一组新的潜在变量; 最后, 似然项 $\log(q_\phi(\mathbf{z}|\mathbf{x}, t, f_\theta(\mathbf{x}, t, \mathbf{z})))$ 尝试对先验和编码器分布进行匹配. 在机器学习文献中, 这种循环匹配过程通常被称为“循环一致性”, 这些约束常以临时方式强制执行. 文献[19-21]表明了循环一致性如何促进从物理空间(通常是高维空间)到潜在空间(相对较低维空间)的一对一非线性投影, 并增强潜在空间的规律性. 这些属性使得这种非线性转换更加稳定和“可解释”[21].

针对式(15)中 $\lambda = 1.0$ 的特殊情况, 所提出的模型类似于传统的有条件生成对抗网络[22]. 本文之所以选择使用反向KL散度来训练网络模型, 主要是因为引入了基于熵的正则化, 用于实现强大且稳定的训练过程, 同时帮助缓解模式崩溃等问题[23]. 当然, 目前许多对抗性模型都有相同的动机, 但是本文所提出方法的一个独特优势在于, 熵正则化过程自然地允许计算关于潜在变量的近似后验分布. 在模型训练中, 本文通过熵正则化推导的公式推断出一个近似的后验分布 $q_\phi(\mathbf{z}|\mathbf{x}, t, \mathbf{u})$, 它本质上回答了这样一个问题, 即在系统输入和输出给定的情况下, 哪些潜在变量最“有可能”出现. 由于输入与输出之间的关系受偏微分方程的约束, 所得到的潜在变量在后验分布上也将受到该物理约束的“启示”. 总体而言, 这些变量旨在总结引入系统随机性的可能高维因素. 结合循环一致性的有利特性, 本文假设分离的潜在变量可以为获得用于非线性模型降阶的良好且与物理相关的特征提供机制.

从理论上讲, 最优参数集 $\{\theta^*, \phi^*, \psi^*\}$ 对应于式(14)和(15)中损失函数定义的两人博弈纳什均衡, 这也间接证明了精确模型分布和潜在变量的精确后验分布可以被恢复[12]. 在实践中, 虽然不能保证达到最优解, 但可以通过交替优化等式(14)和(15)中的两个目标, 并使用随机梯度下降来训练生成模型, 如

$$\max_{\psi} \mathcal{L}_{\mathcal{D}}(\psi), \quad (16)$$

$$\min_{\theta, \phi} \mathcal{L}_{\mathcal{G}}(\theta, \phi). \quad (17)$$

2.3 具有物理信息约束的对抗训练

为了学习式(3)的物理系统概率模型, 可以在生成器的损失函数中适当地引入惩罚项来扩展上述对抗推理框架(见式(15)). 可用的数据对应于散布的观

测集 $\{(\mathbf{x}_i, t_i), \mathbf{u}_i\} (i = 1, 2, \dots, N_u)$, 这些观测数据来自已知的初始或边界条件, 或其他(可能带有噪声的)位置测量数据. 类比于确定性PINN算法设置, 通过定义 $r_\theta(\mathbf{x}, t, \mathbf{z}) := \partial f_\theta(\mathbf{x}, t, \mathbf{z})/\partial t + \mathcal{N}_x f_\theta(\mathbf{x}, t, \mathbf{z})$, 引入一个新的条件概率模型 $p_\theta(r|\mathbf{x}, t, \mathbf{z})$, 其与 $p_\theta(\mathbf{u}|\mathbf{x}, t, \mathbf{z})$ 具有相同的参数. 由于希望生成器 $f_\theta(\mathbf{x}, t, \mathbf{z})$ 生成的每个样本都满足PDE约束, 可以将残差视为确定性变量, 即 $r_\theta(\mathbf{x}, t, \mathbf{z}) = r_\theta(\mathbf{x}, t)$, 并在有限点集 N_r 处通过简单地最小化均方损失来强制实施约束, 即

$$\mathcal{L}_{\text{PDE}}(\theta) = \frac{1}{N_r} \sum_{i=1}^{N_r} \|r_\theta(\mathbf{x}_i, t_i) - r_i\|^2. \quad (18)$$

用于训练方程(3)中的物理学信息网络模型的对抗过程采取如下形式:

$$\begin{aligned} & \max_{\psi} \mathcal{L}_{\mathcal{D}}(\psi), \\ & \min_{\theta, \phi} \mathcal{L}_{\mathcal{G}}(\theta, \phi) + \beta \mathcal{L}_{\text{PDE}}(\theta). \end{aligned} \quad (19)$$

其中可以选择正 β 用于惩罚PDE残差. 对于 $\beta > 0$, 残差损失 $\mathcal{L}_{\text{PDE}}(\theta)$ 作为约束正则化项, 会强制执行给定的物理约束, 从而鼓励生成器 $f_\theta(\mathbf{x}, t, \mathbf{z})$ 生成满足基础偏微分方程的样本. 此外, 该方法还鼓励编码器 $q_\phi(\mathbf{z}|\mathbf{x}, t, f_\theta(\mathbf{x}, t, \mathbf{z}))$ 学习一组与基础物理相关的时空潜在变量 \mathbf{z} , 其可能为复杂系统的概率模型阶数简化开辟新的方向. 图1对上述概率物理学信息网络算法的执行流程进行了总结.

下面给出进行分布信息预测的流程. 完成网络模型训练后, 可以从先验分布 $p(\mathbf{z})$ 中采样潜在变量, 并引入生成器中来生成满足预测模型概率分布 $p_\theta(\mathbf{u}|\mathbf{x}, t, \mathbf{z})$ 下的样本 $\mathbf{u} = f_\theta(\mathbf{x}, t, \mathbf{z})$, 以此构建一个解 $p(\mathbf{u}|\mathbf{x}, t, \mathbf{z})$ 的概率集合. 注意到, 尽管此分布的显式形式未知, 但可以通过蒙特卡罗采样有效地计算其在任何时刻的分布信息. 与训练网络模型的成本相比, 预测步骤的成本可以忽略不计, 因为它仅涉及对生成器 $f_\theta(\mathbf{x}, t, \mathbf{z})$ 进行一次前向传递. 最后, 也是仅需要计算新测试点 (\mathbf{x}^*, t^*) 上的预测分布的均值和方差即可完成对数据分布信息的预测, 有

$$\begin{aligned} \mu_{\mathbf{u}}(\mathbf{x}^*, t^*) &= E_{p_\theta}[\mathbf{u}|\mathbf{x}^*, t^*, \mathbf{z}] \approx \\ & \frac{1}{N_s} \sum_{i=1}^{N_s} f_\theta(\mathbf{x}^*, t^*, \mathbf{z}_i), \end{aligned} \quad (20)$$

$$\begin{aligned} \sigma_{\mathbf{u}}^2(\mathbf{x}^*, t^*) &= \text{Var}_{p_\theta}[\mathbf{u}|\mathbf{x}^*, t^*, \mathbf{z}] \approx \\ & \frac{1}{N_s} \sum_{i=1}^{N_s} [f_\theta(\mathbf{x}^*, t^*, \mathbf{z}_i) - \mu_{\mathbf{u}}(\mathbf{x}^*, t^*)]^2. \end{aligned} \quad (21)$$

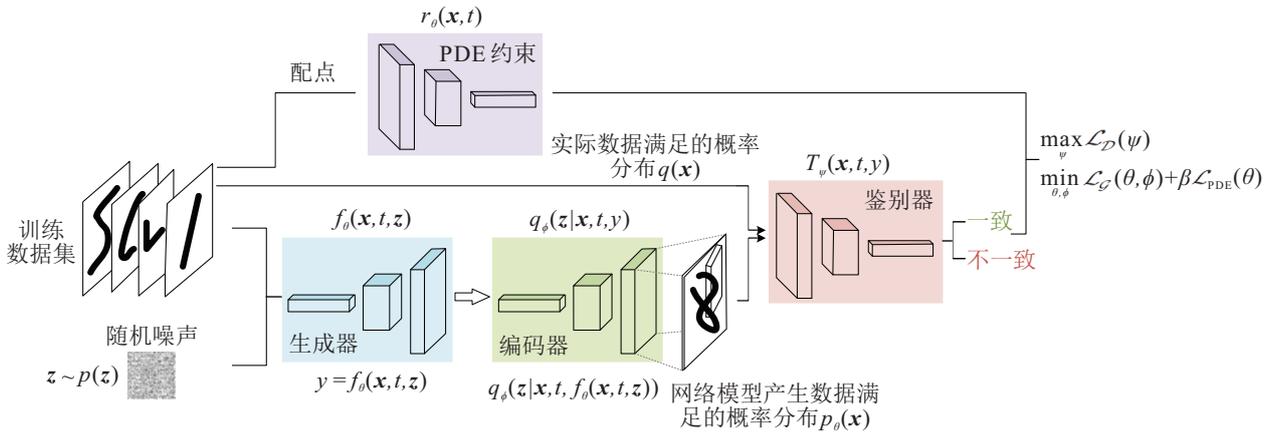


图1 概率物理学信息神经网络算法

其中 $z_i \sim p(z), i = 1, 2, \dots, N_s, N_s$ 对应于蒙特卡罗样本的总数. 在第3.1节展示的示例中, 使用 $N_s = 105$ 个蒙特卡罗本来执行式(20)和(21)的计算, 其可以在1s内完成, 因为它仅涉及已训练好的生成器网络的一次前向传递.

3 算法应用

本文所有例子使用Adam优化器^[24], 训练次数为5000, 学习率为 10^{-4} , 同时固定鉴别器与生成器更新频率比为1:5. 此外, 概率物理学信息神经网络算法默认将熵正则化和残差惩罚参数分别设置为 $\lambda = 1.5$ 和 $\beta = 1$.

3.1 简单验证示例

首先通过一个简单的例子展示PPINN算法的基本功能, 其对应于如下非线性二阶常微分方程:

$$u_{xx} - u^2 u_x = f(x), \quad x \in [-1, 1],$$

$$f(x) = -\pi^2 \sin(\pi x) - \pi \cos(\pi x) \sin^2(\pi x). \quad (22)$$

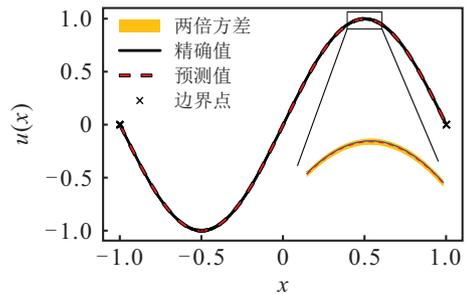
假设该系统具有边界不确定性, 其满足分布 $u(-1), u(1) \sim N(0, \sigma_n^2 \mathbf{I})$. 对于式(22), 相应于 $\sigma_n^2 = 0$ 的确定性解可以容易地得到, 即 $u(x) = \sin(\pi x)$. 给定 N_u 个 $u(x)$ 的观测值, 对应不同的随机边界条件, 并选取 N_r 个配点, 用于计算PDE惩罚项.

此处的目标是通过训练, 获取满足式(22)物理约束的生成网络模型 $u = f_\theta(x, z), z \sim p(z)$. 该模型受式(22)的约束, 以获得解的概率表示 $p_\theta(u|x, z)$. 为此引入3种确定性映射, 即由深度神经网络参数化的 $f_\theta(x, z)$ 、 $q_\phi(x, u)$ 和 $T_\psi(x, u)$, 分别对应于第2.2节介绍的生成器、编码器和鉴别器函数. 此外, 构造一个满足式(22)物理知识约束的神经网络 $r_\theta(x)$, 作用是强制在一组随机分布的点 $x \in [-1, 1]$ 处执行微分方程约束. 所有神经网络都被选择为具有每层50个神经元的两个隐藏层, 并且具有双曲正切激活函数. 此

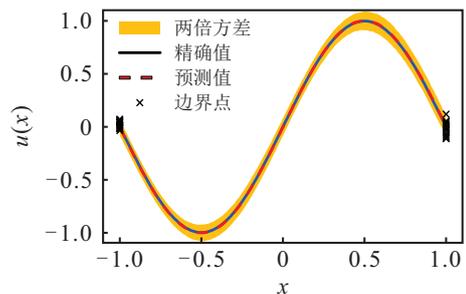
外, 潜在变量的维数设置为1, 即 $z = z$, 且假设 $p(z) \sim \mathcal{N}(0, 1)$. $u(x)$ 的训练数据反映了边界条件的不确定性, 而潜在变量 z 的作用便是使此不确定性传播到由网络模型生成的预测解 $p_\theta(u|x, z)$ 中.

在此考虑了对应于确定性和随机边界条件的两种情况, 即 $\sigma_n^2 = 0$ (确定性数据) 和 $\sigma_n^2 = 0.05$ (5% 高斯不确定性). 在所有情况下, 训练数据包括 $N_u = 20$ 个边界点 $u(-1), u(1)$, 以及用于满足式(22)物理约束的 $N_r = 100$ 个配点, 在 $[-1, 1]$ 内随机选取. 利用概率物理学信息神经网络算法进行训练, 并利用学习出的网络模型进行预测, 结果如图2和图3所示.

具体而言, 图2(a)显示了网络模型预测均值和两个标准偏差, 绘制出此问题的精确解. 注意到, 此情况对应于边界条件的确定性训练数据, 因此精确解



(a) 网络模型在2000次预测下确定系统解平均值和标准差



(b) 网络模型在2000次预测下随机系统解平均值和标准差

图2 确定性和随机边界条件的仿真结果

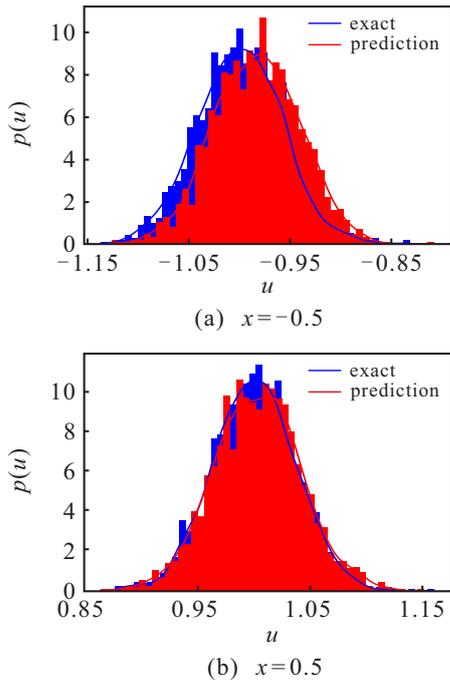


图3 蒙特卡罗解预测边际密度

确定性的. 此处的预测误差在相对 L_2 范数下测量为 $\mathcal{E}_{L_2} = 1.36 \cdot 10^{-3}$, 有

$$\mathcal{E}_{L_2} := \sqrt{\frac{\sum_{i=1}^{N^*} [\mu_u(x_i^*) - u(x_i^*)]^2}{\sum_{i=1}^{N^*} [\mu_u(x_i^*)]^2}}, \quad (23)$$

其中 $N^* = 200$ 表示区间 $[-1, 1]$ 中等间距测试点 x^* 的总数. 此外, 在图2(a)插图中显示的方差量化了与生成模型预测相关的不确定性. 由于 PPINN 方法是基于有限可访问数据的假设, 即使边界上没有噪声这种不确定性仍然存在. 经过后续大量仿真能够观察到, 随着边界上的训练数据量增加, 预测的不确定性会变得更小.

图2(b)显示了随机边界条件对应的预测和不确定性估计结果. 在这种情况下, 生成网络模型的预测不确定性反映了边界条件的随机性和神经网络逼近中固有不确定性的总不确定性. 换句话说, 不清楚结果不确定性的哪一部分是应归因于训练数据的有限大小, 还是系统的随机性/噪声. 这是由于算法假设的 $u = f_\theta(x, t, z)$ 表示中, 潜在变量与输入以复杂的非线性方式纠缠在一起, 虽然能够处理涉及复杂随机过程的情况, 但也阻止了直接量化随机变量与有限数量的训练数据对结果预测不确定性的影响.

学习出的网络模型可以通过条件概率密度 $p_\theta(\mathbf{u}|\mathbf{x}, \mathbf{z})$ 返回解决方案的完整统计特征, 图3提供了预测概率密度 $p_\theta(\mathbf{u}|\mathbf{x}, \mathbf{z})$ 和对应于空间位 $x = -0.5$, $x = 0.5$ 的参考蒙特卡罗解之间一维边缘的视觉比较结果.

由图3结果可知, 所提出的概率物理学信息神经网络算法预测数据的概率分布模型与真实数据概率密度模型能够较好地吻合, 这也验证了所提出算法的有效性.

为了研究概率物理学信息神经网络算法获得结果对熵和 PDE 正则化惩罚参数 λ , β 的敏感性, 针对不同的 λ 和 β 值进行重复模拟, 并计算预测分布 $p_\theta(\mathbf{u}|\mathbf{x}, \mathbf{z})$ 相对于通过蒙特卡罗采样求解器计算的参考“真实”分布 $p_e(\mathbf{u}|\mathbf{x}, \mathbf{z})$ 的反向 KL 散度. 最后在区间 $[-1, 1]$ 的均匀离散化中对反向 KL 散度进行平均, 即

$$E_{p(x)q(z)}\{\text{KL}[p_\theta(u|x, z)||p_e(u|x)]\}. \quad (24)$$

表1展示了 λ 和 β 取不同值后的结果, 揭示了 $\lambda = 1.0$ 时对抗模型模式崩溃的病理性, 以及在 $\beta > 0$ 时引入物理约束条件的优势. 总体而言, 这些结果表明, 当参数 $\lambda > 1.0$ 和 $\beta > 0$ 时, 所提出方法能够返回具有低灵敏度的稳定结果.

表1 不同 λ 和 β 下平均反向 KL 散度

β, λ	1.0	1.5	2.0	5.0
0.0	5.0×10^5	7.5×10^1	6.0×10^1	4.4×10^1
1.0	3.3×10^2	1.8×10^{-1}	2.9×10^{-1}	2.0×10^{-1}
2.0	2.1×10^2	1.7×10^{-1}	5.0×10^{-2}	1.2×10^{-1}
2.0	3.5×10^1	1.8×10^{-1}	1.9×10^{-1}	1.1×10^{-1}

尽管该示例较为简单, 但能够展示出所提 PPINN 算法对非线性偏微分方程在不确定性演化分析方面的基本能力. 与先前从数据推断偏微分方程解的方法^[25]相比, 所提出方法不依赖于高斯假设, 并且可以直接解决非线性问题, 无需进行线性化处理.

3.2 Burgers 方程预测

通过一个更复杂的空间维度中的非线性时变 Burgers 方程对概率物理学信息神经网络算法进行分析, 以进一步验证所提出算法的有效性. Burgers 方程表达式如下:

$$\frac{\partial y}{\partial t} + y \frac{\partial y}{\partial x} - \nu \frac{\partial^2 y}{\partial x^2} = 0,$$

$$x \in [-1, 1], t \in [0, 1],$$

$$y(0, x) = -\sin(\pi x), y(t, -1) = y(t, 1) = 0, \quad (25)$$

其中 $\nu = (0.01/\pi)$ 用于产生一个强非线性响应. 式(25)所示的非线性偏微分方程能够通过科尔霍普夫变换^[26]得到精确解, 这一精确解将用于测试所提出算法生成网络模型的预测有效性.

使用形如 $y = f_\theta(x, t, z)$ 的具有物理信息的网络模型表示未知解 $y(x, t)$. 算法建立3个深度神经网络, 即生成器 $f_\theta(x, t, z)$ 、编码器 $q_\phi(x, t, y)$ 和鉴别器

$T_\psi(x, t, y)$. 其中:生成器和编码器均设置4个隐藏层,每层50个神经元;鉴别器网络设置3个隐藏层,每层50个神经元. 在所有情况下,激活函数均选择双曲正切非线性函数. 潜在变量的先验分布 $p(z)$ 设为一维各向同性高斯分布,即 $z = z, z \sim \mathcal{N}(0, 1)$. 使用 $N_u = 150$ 个输入/输出对训练概率模型,其中50个点为 $y(x, t)$ 的初始条件(根据式(25)中第2个等式选取),另外每个域边界分配50个点(对应式(25)中第3个等式);设置 $N_r = 10\,000$ 个匹配点用于估算式(18)所示的损失,以强制执行Burgers方程偏微分物理约束,所有数据点均在式(25)给出的定义域内随机选择.

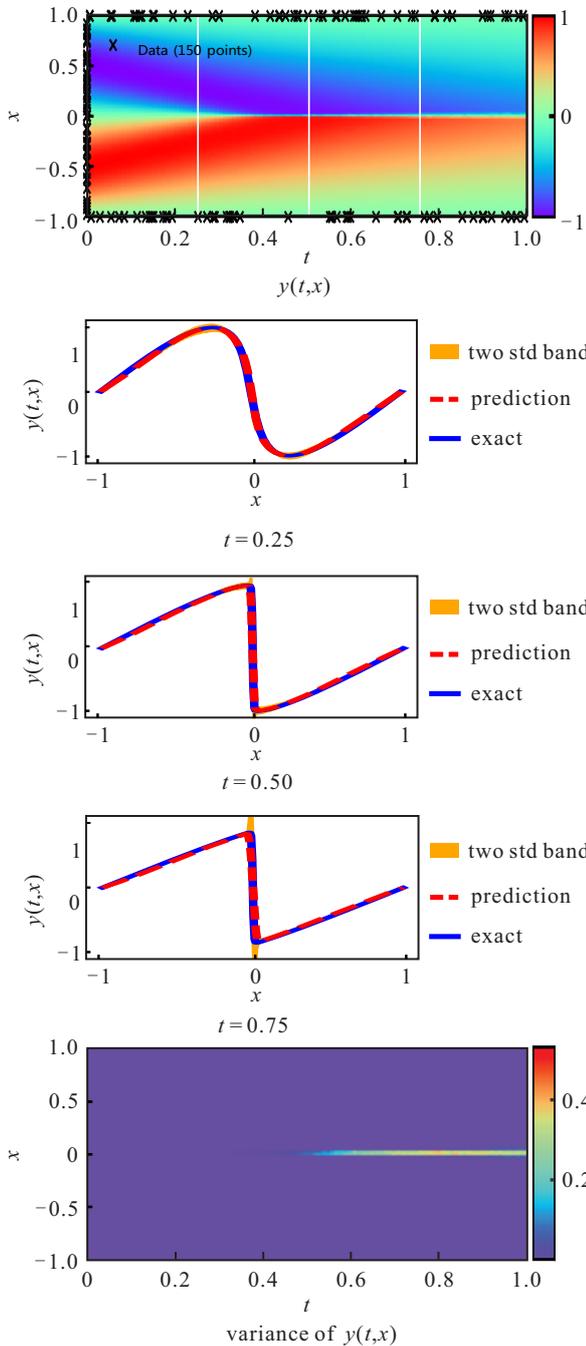


图4 PPINN算法对确定性Burgers方程解的预测

上述所采用的训练数据是无噪声的,因此Burgers的解是确定性的,而该算法生成的概率网络模型 $p_\theta(y|x, t, z)$ 捕获的结果不确定性,提供了有限数量训练数据下生成网络模型逼近程度的误差量化,这在相对 L_2 范数下被测量为 $\varepsilon_{L_2} = 4.1 \times 10^{-2}$. 如文献[14]所讨论的,通过使用拟牛顿优化器可以实现更高的逼近精度. 然而,本文选择使用Adam更新的随机梯度下降,以突显所提出方法在网络模型预测不完全准确时返回不确定性估计的能力,仿真结果如图4所示.

在Burgers方程的初始条件中加入非加性、非高斯的噪声破坏,并在 $x = 0$ 附近加入更大的噪声信

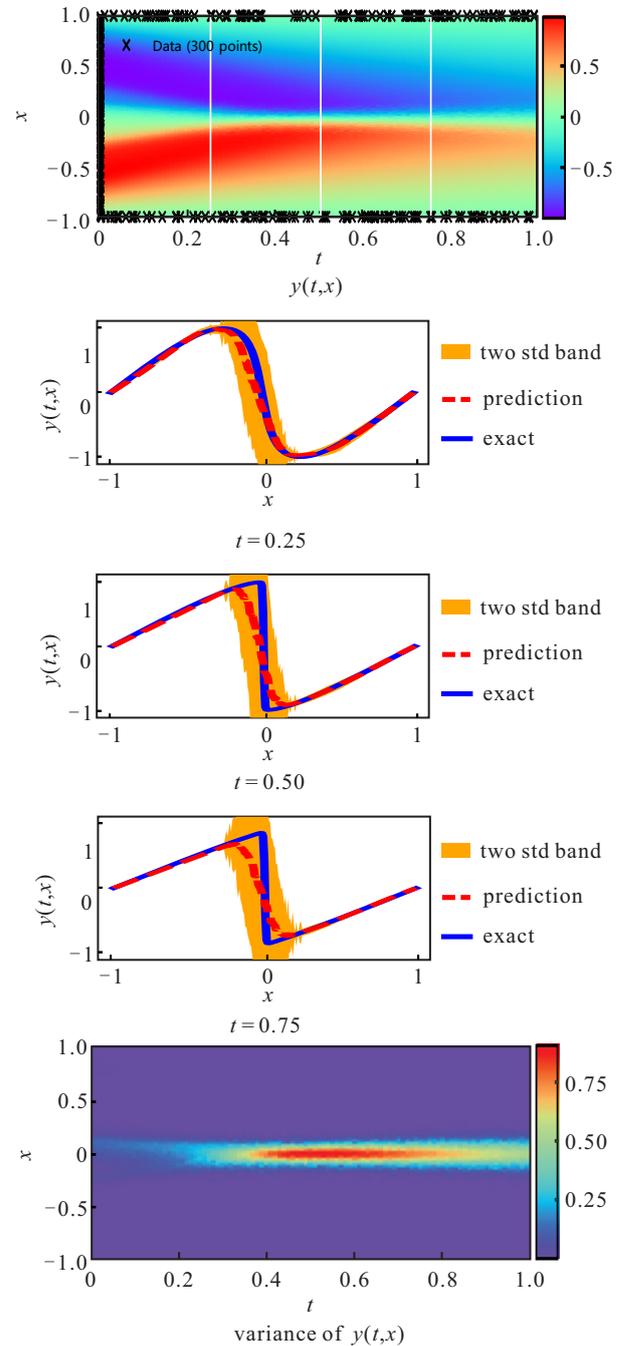


图5 PPINN算法对随机Burgers方程解的预测

号,以放大不确定性对 Burgers 方程在 $x = 0$ 处梯度陡峭的影响. Burgers 方程的随机初始条件为

$$y(x, 0) = -\sin(\pi(x + 2\delta)) + \delta, \\ \delta = \frac{\xi}{e^{3|x|}}, \xi \in \mathcal{N}(0, 0.01). \quad (26)$$

需要说明的是,算法中的神经网络架构以及训练点的数量和位置均与上述确定性分析设置一致.

仿真结果如图 5 所示. 可以看出, PPINN 算法生成的概率网络模型能够有效捕获到随机初始条件在 Burgers 方程中引起的空间和时间解中的不确定性. 正如预期,不确定性集中在 $x = 0$ 处的梯度陡峭区域. 虽然在此仅绘制了部分时域内的 Burgers 方程解情况,但必须强调的是,学习出的概率网络模型 $p_\theta(y|x, t, z)$ 能够提供其非高斯统计的完整概率特性的描述.

4 结 论

本文提出了一个概率物理学信息神经网络算法,能够近似任意条件概率密度,同时被物理偏微分方程所约束. 该算法中还有一个灵活的正则化对抗性推理框架,能够直接从包含随机信息和不完整的测量数据中对网络模型进行训练,系统输入/输出中的不确定性通过一组与底层物理相关的潜在变量进行捕捉. 算法的这些特性使其能够对不确定系统进行概率计算,在小样本状态下训练生成网络模型,处理复杂的噪声过程,并对物理系统进行不确定性演化分析,在系统性能分析、鲁棒轨迹优化以及鲁棒控制方向有很好的应用前景.

尽管所提出的概率物理学信息神经网络算法具有很大的灵活性,但其性能对生成器与鉴别器之间的相互作用较为敏感. 因此,未来的工作将对生成器与鉴别器之间的相互作用进一步研究,以增强所提出算法的鲁棒性.

参考文献(References)

- [1] 王雪松, 荣小龙, 程玉虎, 等. 基于自适应多尺度图卷积网络的多标签图像识别[J]. 控制与决策, 2022, 37(7): 1737-1744.
(Wang X S, Rong X L, Cheng Y H, et al. Multi-label image recognition based on adaptive multi-scale graph convolutional network[J]. Control and Decision, 2022, 37(7): 1737-1744.)
- [2] Raissi M, Perdikaris P, Karniadakis G E. Physics informed deep learning (part I): Data-driven solutions of nonlinear partial differential equations[J/OL]. 2017, arXiv: 1711.10561.
- [3] Raissi M, Perdikaris P, Karniadakis G E. Physics informed deep learning (part II): Data-driven discovery of nonlinear partial differential equations[J/OL]. 2017, arXiv: 1711.10566.
- [4] Lagaris I E, Likas A, Fotiadis D I. Artificial neural networks for solving ordinary and partial differential equations[J]. IEEE Transactions on Neural Networks, 1998, 9(5): 987-1000.
- [5] Oden J T, Moser R, Ghattas O. Computer predictions with quantified uncertainty, partII[J]. SIAM News, 2010, 43(10): 2008-2011.
- [6] 张保强, 陈梅玲, 孙东阳, 等. 基于概率盒演化的时变系统不确定性量化方法[J]. 控制与决策, 2020, 35(10): 2459-2465.
(Zhang B Q, Chen M L, Sun D Y, et al. Uncertainty quantification for time-variant system based on probability box evolution[J]. Control and Decision, 2020, 35(10): 2459-2465.)
- [7] Gao B B, Hu G G, Zhong Y M, et al. Distributed state fusion using sparse-grid quadrature filter with application to INS/CNS/GNSS integration[J]. IEEE Sensors Journal, 2022, 22(4): 3430-3441.
- [8] 邵天浩, 张宏军, 程恺, 等. 抽象技术及其在蒙特卡洛树搜索中的应用研究综述[J]. 控制与决策, 2024, 39(4): 1075-1094.
(Shao T H, Zhang H J, Cheng K, et al. A review of abstract technology and its application in Monte Carlo tree search[J]. Control and Decision, 2024, 39(4): 1075-1094.)
- [9] 王子赞, 程林, 王艳, 等. 基于正交超平空间定向扩展的滤波故障诊断方法[J]. 控制与决策, 2022, 37(12): 3223-3232.
(Wang Z Y, Cheng L, Wang Y, et al. Orthometric hyperparallel spatial directional expansion filtering based fault diagnosis method[J]. Control and Decision, 2022, 37(12): 3223-3232.)
- [10] 肖文鑫, 张文文. 一种基于概率关联的局部高斯过程回归算法[J]. 自动化学报, 2022, 48(8): 1940-1949.
(Xiao W X, Zhang W W. A local Gaussian regression algorithm based on probability association[J]. Acta Automatica Sinica, 2022, 48(8): 1940-1949.)
- [11] 左为恒, 宋璐璐. 基于改进堆叠自动编码器的循环冷却水系统工艺介质温度预测控制方法[J]. 控制与决策, 2020, 35(12): 2835-2844.
(Zuo W H, Song L L. Predictive control method of process medium temperature in circulating cooling water system based on improved stacked auto encoders[J]. Control and Decision, 2020, 35(12): 2835-2844.)
- [12] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J/OL]. 2014, arXiv: 1406.2661.

- [13] 梁相龙, 姚建勇. 基于神经网络的机电伺服系统非线性控制[J]. 控制与决策, 2023, 38(4): 1008-1014.
(Liang X L, Yao J Y. Nonlinear control of mechatronic servo system based on neural network[J]. Control and Decision, 2023, 38(4): 1008-1014.)
- [14] Raissi M, Perdikaris P, Karniadakis G E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. Journal of Computational Physics, 2019, 378: 686-707.
- [15] Ghanem H, Salmon J, Keriven N, et al. Supervised learning of analysis-sparsity priors with automatic differentiation[J]. IEEE Signal Processing Letters, 2023, 30: 339-343.
- [16] Puranik B, Madhow U, Pedarsani R. Generalized likelihood ratio test for adversarially robust hypothesis testing[J]. IEEE Transactions on Signal Processing, 2022, 70: 4124-4139.
- [17] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs[C]. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, 2016: 2234-2242.
- [18] 李小南, 赵璐, 易黄建. 基于加权信息熵的直觉模糊信息系统的三支决策[J]. 控制与决策, 2022, 37(10): 2705-2713.
(Li X N, Zhao L, Yi H J. Three-way decision of intuitionistic fuzzy information systems based on the weighted information entropy[J]. Control and Decision, 2022, 37(10): 2705-2713.)
- [19] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. IEEE International Conference on Computer Vision. Venice, 2017: 2242-2251.
- [20] Lee H Y, Tseng H Y, Huang J B, et al. Diverse image-to-image translation via disentangled representations[C]. European Conference on Computer Vision. Cham: Springer, 2018: 36-52.
- [21] Jha A H, Anand S, Singh M, et al. Disentangling factors of variation with cycle-consistent variational auto-encoders[C]. European Conference on Computer Vision. Cham: Springer, 2018: 829-845.
- [22] 孔锐, 蔡佳纯, 黄钢, 等. 基于复合生成对抗网络的对抗样本生成算法研究[J]. 控制与决策, 2023, 38(2): 528-536.
(Kong R, Cai J C, Huang G, et al. Research on generative adversarial example algorithm based on multiple GANs[J]. Control and Decision, 2023, 38(2): 528-536.)
- [23] Yang L, Zhang D K, Karniadakis G E. Physics-informed generative adversarial networks for stochastic differential equations[J/OL]. 2018, arXiv: 1811.02033.
- [24] Kingma D P, Ba J. Adam: A method for stochastic optimization[J/OL]. 2014, arXiv: 1412.6980.
- [25] Raissi M, Perdikaris P, Karniadakis G E. Numerical Gaussian processes for time-dependent and nonlinear partial differential equations[J]. SIAM Journal on Scientific Computing, 2018, 40(1): A172-A198.
- [26] Hopf E. The partial differential equation $u_t + u_{xx} = u_{xx}$ [J]. Communications on Pure and Applied Mathematics, 1950, 3(3): 201-230.

作者简介

曹瑞(1994—), 女, 讲师, 博士, 从事导航、制导与控制的研究, E-mail: stdio@yzu.edu.cn;

裔扬(1980—), 男, 教授, 博士, 从事制科学与工程的研究, E-mail: yiyang@yzu.edu.cn;

刘燕斌(1980—), 男, 教授, 博士, 从事导航、制导与控制的研究, E-mail: nuaa_liuyanbin@139.com.