

物理信息神经网络

—用于解决涉及非线性偏微分方程的正反问题的深度学习框架

Abstract: 我们介绍了具有物理知识的神经网络-经过训练的神经网络，可以解决监督学习任务，同时遵守通用非线性偏微分方程描述的任何给定的物理定律。在这项工作中，我们在解决两类主要问题的背景下介绍了我们的发展：数据驱动的解决方案和偏微分方程的数据驱动的发现。根据可用数据的性质和安排，我们设计了两种不同类型的算法，即连续时间模型和离散时间模型。前一种类型一个新的数据高效时空函数逼近器族，而后一种类型则允许使用任意准确的隐式 Runge-Kutta 时间步长方案，且步数不受限制。提出的框架的有效性通过流体、量子力学、反应扩散系统以及非线性浅水波传播方面的一系列经典问题得到证明。

Keywords: 数据驱动的科学计算, 机器学习, 预测模型, 龙格库塔方法, 非线性动力学.

1. Introduction

随着可用数据和计算资源的爆炸性增长，机器学习和数据分析的最新进展已在包括图像识别 [1]，认知科学 [2] 和基因组学 [3]。但是，在分析复杂的物理，生物或工程系统的过程中，数据获取的成本往往是高得令人望而却步，而且我们不可避免地面临根据部分信息得出结论和做出决策的挑战。在这种小数据体制下，绝大多数最新的机器学习技术（例如深度/卷积/递归神经网络）都缺乏鲁棒性，无法提供任何收敛保证。

乍一看，训练一个深度学习算法从几个（可能是非常高维的）输入和输出数据对中准确地识别非线性映射的任务似乎是最幼稚的。对于许多与物理和生物系统建模有关的案例来说，有大量的先验知识目前还没有被用于现代机器学习实践中。让它成为支配系统的时变动力学的原则物理定律，或一些经验验证的规则或其他领域的专业知识，这些先验信息可以作为一个正则化代理，将可接受的解决方案的空间限制到一个可管理的大小。（例如，在不可压缩流体动力学问题中，丢弃任何违反质量守恒原理的不现实的流动解）。作为回报，将这种结构化信息编码为学习算法会使算法看到的数据的信息内容得到放大，从而在即使只有几个训练示例可用的情况下，也能使其迅速转向正确的解决方案并获得良好的解。

利用结构化的先验信息来构建高效的数据和物理信息学习机器的前景的第一次瞥见已经在最近的研究 [4-6] 中得到展示。在那里，作者使用高斯过程回归 [7] 来设计函数表示，这是量身定制的一个给定的线性算子，并能够准确地推断解决方案和提供不确定估计的几个原型问题的数学物理。在随后的研究中，Raissi et al 提出了非线性问题的扩展。[8,9] 在推理和系统识别的上下文中。尽管高斯过程在编码先验信息方面具有灵活性和数学上的优雅性，但非线性问题的处理引入了两个重要的限制。首先，在 [8,9] 中，作者不得不在时间上局部线性化任何非线性项，从而限制了所提方法在离散时间域的适用性，并损害了其在强非线性条件下的预测准确性。。其次，高斯过程回归的贝叶斯特性要求一定的先验假设，这可能会限制模型的表示能力，导致鲁棒性问题，特别是对于非线性 [10] 问题。

2. Problem setup

在这项工作中，我们采用一种不同的方法，使用深度神经网络和利用其众所周知的能力作为通用函数逼近 [11]。在这种情况下，我们可以直接处理非线性问题，而不需要遵守任何预先假设、线性化或局部时间逼近。我们利用最近在自动微分 [12] 的发展-最有用但可能在科学计算中未充分利用的技术之一根据输入坐标和模型参数对神经网络进行微分，从而获得具有物理信息的神经网络。此类神经网络必

须遵守由一般时变和非线性偏微分方程建模的源自控制观察数据的物理定律的任何对称性，不变性或守恒原理。这种简单而强大的结构使我们能够解决计算科学中的各种问题，并引入了一种潜在的变革性技术，从而导致了一系列各种的发展。[发展](#) (Figura 1).

这项工作的总体目标是为建模和计算的新范式奠定基础，从而丰富数学物理长期发展的深度学习。为此，我们的手稿被分为两部分，目的是呈现我们在两大类问题的背景下的发展：偏微分方程的数据驱动的解决方案和数据驱动地发现。该手稿的所有代码和数据集都可以在 GitHub 上下载：<https://github.com/maziarraissi/PINNs>。在这项工作中，我们一直使用相对简单的具有双曲正切激活函数的深度前馈神经网络结构，并且没有额外的正则化（例如 L1/L2 惩罚、dropout 等）。手稿中的每个数值例子都伴随着一个关于我们所使用的神经网络架构的详细讨论，以及关于它的训练过程的细节（例如优化器，学习率等）。最后，附录 a 和附录 b 提供了旨在证明所提方法性能的一系列全面系统的研究。

在这项工作中，我们考虑一般形式的参数化和非线性偏微分方程：

$$u_t + N[u; \lambda] = 0, x \in \Omega, t \in [0, T] \quad (1)$$

该方程封装了数学物理学中广泛的问题，包括守恒定律、扩散过程、对流-扩散-反应系统和动力学方程。下标表示在时间或空间上的偏微分。给定系统的噪声测量值，我们感兴趣的是解决两个不同的问题。第一个问题是偏微分方程 [4,8] 的推论，滤波和平滑或数据驱动的解决方案，这些问题指出：给定固定模型参数 λ ，关于未知未知状态 $u(t, x)$ 可以说什么。第二个问题是偏微分方程的学习、系统识别或数据驱动地发现 [5,9,14]，陈述：最能描述观测数据的参数 λ 是什么？

3. Data-driven solutions of partial differential equations

让我们首先集中讨论计算一般形式的偏微分方程的数据驱动解的问题（即上面概述的第一个问题）：

$$u_t + N[u] = 0, x \in \Omega, t \in [0, T], \quad (2)$$

我们提出了两种不同类型的算法，即连续时间模型和离散时间模型，并通过不同的基准问题突出它们的性质和性能。在我们研究的第二部分（见第 4 部分），我们将注意力转移到数据驱动的偏微分方程发现问题上 [5,9,14]。

3.1. continuous time models

尽管因为微分算子的不同有不同的激活函数，但新定义的 f 与原始解 u 共享网络中的参数。神经网络是由微分算子中的链式法则生成的。共享参数通过最小化均方误差来训练。

尽管在先前的研究中已经探索了使用物理定律来约束神经网络的类似想法 [15,16]，但在这里我们使用现代计算工具对其进行了重新研究，并将其应用于由时变非线性偏微分方程描述的更具挑战性的动力学问题。

在这里，我们应该强调这条工作线与文献中详细阐述在计算物理学中使用机器学习的现有方法之间的重要区别。最近，王等人也使用了“基于物理的机器学习”这个术语。湍流建模背景下的 [17]。其他用于物理系统预测建模的机器学习方法的例子包括 [18-29]。所有这些方法都将机器学习算法如支持向量机、随机森林、高斯过程和前馈/卷积/循环神经网络作为黑盒工具。如上所述，这项工作的目标是更进一步，通过重新构建“定制”激活和损失函数，这些函数是为潜在的微分算子量身定制的。这让我们能够通过理解和欣赏深度学习领域内自动分化所扮演的关键角色来打开黑箱。一般来说，自动微分，特别是反向传播算法，是目前训练深度模型的主要方法，通过对模型的参数（例如，权值和偏差）进行求导。在这里，我们

使用深度学习社区使用的完全相同的自动微分技术，通过对输入坐标（即空间和时间）的导数，向神经网络提供物理信息，其中物理是由偏微分方程描述的。我们经验地观察到，这种结构化的方法引入了一种正则化机制，使我们能够使用相对简单的前馈神经网络结构，并用少量的数据对其进行训练。这一简单想法的有效性可能与 Lin、Tegmark 和 Rolnick[30] 的评论有关，并提出了许多有趣的问题，需要在未来的研究中定量解决。为此，拟议的作品从 Psychogios 和 Ungar [16]，Lagarisetal 的早期贡献中获得灵感。以及 Kondor[31,32]、Hirn[33] 和 Mallat[34] 的当代作品。

在所有用于说明偏微分方程的数据驱动解决方案的情况下，训练数据 N_u 的总数相对较小（几百到几千个点），我们选择使用 L-BFGS 优化所有损失函数，这是一种准牛顿、全批梯度的 [35] 优化算法。对于更大的数据集，如第 4 节中讨论的数据驱动模型发现示例，可以使用随机梯度下降及其现代变体轻松使用计算效率更高的小型批处理设置 [36,37]。尽管没有理论上的保证，这个过程收敛于全局最小值，我们的经验证据表明，如果给定的偏微分方程是适定的，其解决方案是独一无二的，我们的方法可以实现良好的预测精度（条件是有充分表达神经网络架构和 N_f 搭配足够数量的点）。这一一般性观察结果与由方程 (4) 的均方误差损失所导致的优化景观密切相关，并定义了一个与深度学习的最新理论发展同步有待研究的开放问题 [38,39]。为此，我们将使用附录 a 和附录 b 中提供的一系列系统敏感性研究来测试所提出方法的稳健性。

3.1.1. Schrodinger equation

此示例旨在强调我们的方法处理周期性边界条件，复值解以及控制偏微分方程中不同类型的非线性的能力。一维非线性 Schrödinger 方程是一个经典的场方程，用于研究量子力学系统，包括光纤波导中的非线性波传播，玻色-爱因斯坦凝聚物和等离子体波。在光学中，非线性项源自给定材料的强度相关折射率。同样，玻色-爱因斯坦凝聚物的非线性项是相互作用的 N 体系统的平均场相互作用的结果。非线性 Schrödinger 方程以及周期边界条件由下式给出：

$$\begin{cases} ih_t + 0.5h_{xx} + |h|^2h = 0, & x \in [-5, 5], \quad t \in [0, \pi/2] \\ h(0, x) = 2 \operatorname{sech}(x) \\ h(t, -5) = h(t, 5) \\ h_x(t, -5) = h_x(t, 5) \end{cases}$$

$$f := ih_t + 0.5h_{xx} + |h|^2h$$

为了评估我们方法的准确性，我们使用传统的谱方法模拟了公式 (5)，以创建高分辨率数据集。具体而言，从初始状态 $h(0, x) = 2 \operatorname{sech}(x)$ 开始，并假设周期性边界条件 $h(t, -5) = h(t, 5)$ 和 $h_x(t, -5) = h_x(t, 5)$ ，我们使用 Chebfun 软件包 [40] 结合了方程式 (5)，直到最后一个时间 $t = \pi/2$ 为止，该软件包具有 256 个模式的频谱傅立叶离散化和具有时间-的四阶显式 Runge-Kutta 时间积分器步骤 $t \pi/2 \cdot 10^{-6}$ 。在我们的数据驱动设置下，我们观察到的只是在时间 $t = 0$ 时对潜函数 $h(t, x)$ 的测量值 $x_0^i, h_{0i=1}^{i=N_0}$ 。特别是，训练集包括从完整的高分辨率数据集中随机解析的 $h(0, x)$ 上的 $N_0 = 50$ 个数据点，以及用于执行周期边界的 $N_b = 50$ 个随机采样的配置点 $t_{bi}^{N_b} = 1$ 。此外，我们假设 $N_f = 20,000$ 个随机采样的配置点，用于在解决方案域内实施公式 (5)。所有随机采样的点位置都是使用填充拉丁的超立方体采样策略生成的 [41]。

在这里，我们的目标是推断 Schrödinger 方程 (5) 的整个时空解 $h(t, x)$ 。我们选择使用每层 100 个神经元和双曲正切激活函数的 5 层深度神经网络来联合表示潜函数 $h(t, x) = [u(t, x) \quad v(t, x)]$ 。通常，应该为神经网络提供足够的近似能力，以适应预期的 $u(t, x)$ 复杂度。尽管可以使用诸如贝叶斯优化 [42] 之类的更系统

的程序来微调神经网络的设计，但是在没有理论误差/收敛估计的情况下，神经体系结构/训练程序与基本的微分方程的复杂性之间的相互作用仍然知之甚少。通过采用贝叶斯方法并监视预测后验分布的方差，可以找到一种评估预测解决方案准确性的可行途径，但这超出了当前工作的范围，将在以后的研究中进行研究。

在此示例中，我们的设置旨在针对已知的过度拟合问题突出提出的方法的鲁棒性。具体而言，方程式 (6) 中的 MSE_f 中的项用作惩罚不满足方程式 (5) 的解决方案的正则化机制。因此，物理信息神经网络的一个关键特性是可以使用小数据集对其进行有效训练。在物理系统研究中经常遇到的一种环境，对于这种环境而言，数据获取的成本可能会令人望而却步。图 1 总结了我们的实验结果。具体地，图 1 的顶部面板展示出了预测的时空解 $|h(t, x)| = \sqrt{u^2(t, x) + v^2(t, x)}$ 的大小以及初始训练数据和边界训练数据的位置。针对这个问题的测试数据验证了所产生的预测误差，并在相对 L2 范数中测得的预测误差为 $1.97 \cdot 10^{-3}$ 。在图 1 的底部面板中提供了对预测解的更详细的评估。特别是，我们给出了在不同时刻 $t=0.59$ 、 0.79 、 0.98 的精确解和预测解之间的比较。只需使用少量初始数据，即可通过物理知悉的神经网络准确捕获 Schrödinger 方程的复杂非线性行为。

到目前为止，考虑到的连续时间神经网络模型的一个潜在局限性在于，需要使用大量的搭配点 N_f ，以便在整个时空域中实施物理信息约束。尽管这对于一两个空间维度上的问题不构成重大问题，但它可能会在更高维度的问题上引入严重的瓶颈，因为全局实施物理信息约束所需的配置点总数（即，在我们的情况下是部分约束）微分方程）将成倍增加。尽管可以使用稀疏网格或准蒙特卡洛采样方案在某种程度上解决此限制 [43,44]，但在下一节中，我们将通过引入结构更复杂的神经网络提出另一种结构化的神经网络即利用经典的 Runge-Kutta 时间步长方案 [45] 来避开对匹配点的需求的方法。

3.2. Discrete time models

对方程 (2) 使用经典的 q 阶龙哥库塔方法离散得到：

$$u^{n+c_i} = u^n - \Delta t \sum_{j=1}^q a_{ij} \mathcal{N} [u^{n+c_j}], i = 1, \dots, q$$

$$u^{n+1} = u^n - \Delta t \sum_{j=1}^q b_j \mathcal{N} [u^{n+c_j}]$$

4. Data-drive discovery of partial differential equations

4.1. continuous time models

4.2. Discrete time models

O número de tabelas e figuras utilizadas no artigo deve ser limitado a compreensão e elucidação do texto. Devem ser inseridas no corpo do texto, para identificação da sua posição e do tamanho aproximado (Figura 1).

Tabelas e figuras possuem numeração independente, que deve ser feita sequencialmente na ordem em que são citadas no texto. Devem também ter uma legenda auto-explicativa, sendo que as tabelas terão legendas na parte superior e as figuras as terão na parte inferior, centralizadas em relação à tabela ou figura. A legenda inicia com o termo "Tabela" ou "Figura" (primeira letra em maiúscula), de acordo com o caso, seguido de um espaço e do número de ordem sequencial, em algarismos arábicos, seguido de hífen entre espaços e do texto da legenda, com a primeira letra da primeira palavra em maiúscula e as demais em minúscula, exceção feitas àquelas que normalmente são escritas em maiúsculas. Devem ser citadas no texto como "Tabela 1" e "Figura" seguidas de espaço e do número correspondente.



Figura 1. A typical figure

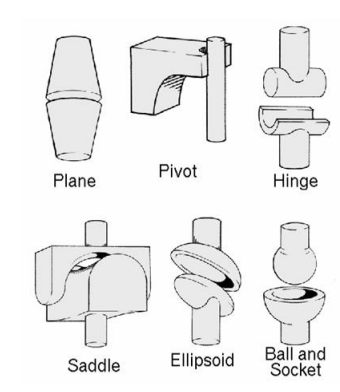


Figura 2. This figure is an example of a figure caption taking more than one line and justified considering margins mentioned in Section ??.

5. Agradecimentos

Usados apenas quando necessário. Devem ser sucintos, diretos e dirigidos apenas a pessoas ou instituições que contribuíram substancialmente para a elaboração do trabalho.

6. Referências

As referências a autores e fontes são inseridas no texto colocando entre parênteses o sobrenome do autor, com inicial maiúscula, seguida da data, conforme o exemplo: (Smith e Jones, 1999). Quando for conhecida a paginação, pode-se incluí-la, por exemplo, (Tarouco, 1991, p. 237). Havendo mais de um título do(s) autor(es) no mesmo ano, deve-se distingui-las utilizando uma letra minúscula (a,b,c) depois da data, por exemplo, (Tarouco, 1991b). Quando houver três ou mais autores, no texto será citado apenas o primeiro autor seguido de "et al.", mas nas referências bibliográficas, no final do artigo, os demais nomes também deverão aparecer. Quando o nome do autor é citado diretamente no texto pode-se colocar entre parênteses apenas a data de publicação, por exemplo,

Boulic e Renault (1991). Na citação de citação, identifica-se a obra diretamente consultada; o autor e/ou a obra citada nesta é assim indicado: (Knuth, 1984) citada por Vicari (Cruz et al., 1998).

As referências completas de cada autor e fonte citadas no texto devem

Tabela 1. Variables to be considered on the evaluation of interaction techniques

	Chessboard top view	Chessboard perspective view
Selection with side movements	6.02 \pm 5.22	7.01 \pm 6.84
Selection with in-depth movements	6.29 \pm 4.99	12.22 \pm 11.33
Manipulation with side movements	4.66 \pm 4.94	3.47 \pm 2.20
Manipulation with in-depth movements	5.71 \pm 4.55	5.37 \pm 3.28

aparecer no final do artigo sob o título "Referências Bibliográficas", ordenadas alfabeticamente pelo(s) sobrenome(s) do(s) autore(s).

Os autores do artigo devem certificar-se de que as referências citadas no texto contam da lista de referências com datas exatas e nomes de autores corretamente grafados. A exatidão dessas referências é de responsabilidade dos autores do artigo. Comunicações pessoais, trabalhos inéditos ou em andamento poderão ser citados quando absolutamente necessários, mas não devem ser incluídos na lista de referências bibliográficas; apenas citados no texto ou nas Notas do texto.

Referências

- Boulic, R.; Renault, O. 3d hierarchies for animation. In: Magnenat-Thalmann, N.; Thalmann, D. (Ed.). New Trends in Animation and Visualization. [S.l.]: John Wiley & Sons Ltd., 1991.
- Cruz, X. Y. Z.; Correa, X. Y. Z.; Costa, X. Y. Z.; Costa, X. Y. Z. Exemplo de citação no texto. [S.l.: s.n.], 1998.
- Knuth, D. E. The \TeX Book. 15th. ed. [S.l.]: Addison-Wesley, 1984.
- Smith, A.; Jones, B. On the complexity of computing. In: Smith-Jones, A. B. (Ed.). Advances in Computer Science. [S.l.]: Publishing Press, 1999. p. 555–566.