arXiv:2302.01810v1 [cs.LG] 3 Feb 2023

使用双目标优化训练 PINN：数据损失与残差损失之间的权衡

Fabian Heldmann[b]、Sarah Berkhahn[a]、Matthias Ehrhardt[a]、Kathrin Klamroth[b]

[a] *University of Wuppertal, Chair of Applied and Computational Mathematics, Gaußstrasse 20, 42119 Wuppertal, Germany*[b] *University of Wuppertal, Chair of Optimization, Gaußstrasse 20, 42119 Wuppertal,Germany*

摘要

物理信息神经网络（PINN）已被证明是一种有效的工具，可用于表示具有测量数据且预期数据动态遵循某些物理定律的问题。

本文提出以多目标视角来审视物理信息神经网络的训练，在真正的双目标优化方法中，将数据损失和残差损失视为两个独立的目标函数。

作为示例，我们研究德国新冠肺炎疫情的预测，并构建了一个扩展的易感者-感染者-康复者（SIR）模型，进一步纳入了易感者、已接种者和住院人群（SVIHR模型），以模拟转移率并预测未来感染人数。SIR类模型由常微分方程组（ODEs）表示。我们研究所生成的物理信息神经网络用于新冠肺炎疫情预测的适用性，并将所得预测曲线与采用非标准有限差分法求解常微分方程组及初始数据所得到的曲线进行比较。

该方法适用于各种定义动力学状态的常微分方程组。这些动力学状态不必由 SIR 型模型描述，相应的基础数据集也不必与新冠肺炎相关。

Email addresses: heldmann@uni-wuppertal.de (FabianHeldmann), streibert@uni-wuppertal.de (Sarah Berkhahn), ehrhardt@uni-wuppertal.de (Matthias Ehrhardt), klamroth@uni-wuppertal.de (KathrinKlamroth)

关键词：物理信息神经网络、舱室模型、损失函数、多目标优化、加权参数、帕累托前沿

1. 引言

物理信息神经网络（PINN） [1, 2], 也称为理论启发的机器学习 [3], ，近年来已成为求解微分方程的一种热门方法。通过将微分方程的残差纳入基于神经网络的替代模型的损失函数，物理信息神经网络能够将测量数据与微分方程所给出的物理约束无缝结合。物理信息神经网络也可视为一种替代模型，通过引入额外数据来求解微分方程；也可视为对底层物理系统进行数据驱动校正（甚至发现）的方法。

截至2022年年底，我们已经历了数波新冠肺炎大流行，不同病毒变异株在不同时期占据主导地位。为遏制疾病不受控制的传播，人们实施了不同程度的干预措施和防护措施。本文以德国新冠肺炎疫情为例，重点关注截至第四波疫情（即奥密克戎变异株疫情）的时期；这波疫情在2022年2月和3月达到峰值。

SARS-CoV-2 的 B.1.617.2（德尔塔）变异株传染性高于此前的 B.1.1.7（阿尔法）、B.1.351（贝塔）和 P.1（伽马）变异株，自 2021年3月 起便已在德国出现，并在2021年数月间成为德国的主要变异株[4]。2021年秋季，新的奥密克戎变体被发现，并被世界卫生组织列为需要关注的变体。截至 11 月底，奥密克戎变体已被划分为三个亚谱系（BA.1、BA.2、BA.3），且相较于德尔塔变异株具有传播优势 2021 [4]。奥密克戎变体迅速传播至全球。奥密克戎变体与德尔塔变异株的三种重组株（XD、XE、XF）已登记为亚谱系 [4]。

德国新冠肺炎第四波疫情期间的局部峰值分别出现在11月 28th 2021 ，为693；2月 14th 2022 ，为2,434；以及3月 20th

2022年，每百万人每日感染人数达到2,619人。这些峰值均高于此前三波疫情的全局峰值：2020年春季（4月 2nd 2020为69人）、2020/2021年冬季（12月 22nd 2020 为 305 人）以及 2021 年春季（ 4 月 25th 2021 为 257 人） [5]。

本研究用于描述新冠肺炎人口动态的数学模型是易感者-已接种者-感染者-住院者-康复者（SVIHR）模型。在该模型中，症状前期个体与有症状者合并在感染者舱室中，因此我们只有一个由未住院人员组成的感染者舱室。

我们的工作主要包括两大部分：首先，我们建立了SVIHR模型，并以此为基础构建物理信息神经网络（PINN）。PINN方法将某些物理信息约束（例如以微分方程表示的约束）纳入相应深度神经网络的损失函数中。因此，常微分方程组在神经网络的训练（即优化）过程中起着至关重要的作用。

Raissi等人首次在其研究中提出了物理信息神经网络，此后该方法被用于求解不同的正问题和逆问题 [1]。PINN方法在训练过程中权衡基于数据的损失函数与物理损失函数。这使搜索合理解的过程倾向于那些在一定程度上满足“物理定律”的解；在本例中，这一物理定律即为SVIHR舱室模型。PINN的损失函数基于近期感染事件的报告数据（数据损失部分），以及包含状态转移和传播动力学的常微分方程组；利用自动微分计算由此得到的所谓残差网络（残差损失部分）。数据损失使物理信息神经网络成为一种数据驱动技术。这里，我们区分两类训练数据：一类覆盖德国疫情暴发以来的整个时期（长期预测），另一类仅涉及一个疫情峰值（短期预测）。

物理信息神经网络涉及若干固定的模型参数，由此计算出转移参数。我们方法的一个独特之处在于，使用非标准有限差分（NSFD）方法估计对系统动力学至关重要的参数。换言之，我们将物理信息神经网络与

用于常微分方程组的数值方法相结合，以便更好地估计残差损失函数中的关键参数。虽然非标准有限差分解通常无法很好地近似数据，但可用于更准确地估计模型参数，例如疫苗接种前后的传播率。

我们工作的第二项贡献是优化数据损失部分与残差损失部分之间关系的加权参数。在物理模型只能部分表示测量数据的情况下，例如预测 COVID-19感染率时，数据损失与残差损失确实存在冲突。虽然主要目标是尽可能准确地再现测量数据，但残差损失更像是一项正则化项，有助于克服数据中的噪声和离群值，并更好地预测潜在动力学。为这两个训练目标选择合理的权重绝非易事。为此，我们将训练过程视为一个双目标优化问题，其中残差损失和数据损失被视为两个相互独立的目标函数。我们不采用预先确定且固定的加权参数来合并这两个目标，而是通过生成帕累托前沿的（粗略）近似来确定合适的加权参数。

对于训练过程，我们采用基于标量化的方法，将双目标问题转换为一系列加权和标量化问题。通过采用自适应选取的标量化参数反复进行训练运行，可以找到较优解。由此得到的帕累托前沿近似结果，提供了有关数据损失与残差损失之间权衡的宝贵信息。一方面，这些信息可用于评估所采用的物理模型是否适用；另一方面，全面分析（近似的）帕累托前沿，有助于做出明智的选择，确定适当的折中方案：根据应用背景以及决策者的偏好和信念，侧重于数据或物理模型。

1.1. 相关研究

自COVID-19大流行暴发以来，人们提出了各种改进的易感-感染-康复（SIR）隔室模型，即舱室模型，用以研究SARS-CoV-2传播的不同方面。物理信息神经网络也已应用于舱室模型，并在COVID-19大流行背景下得到研究。

例如，Malinzi 等人将物理信息神经网络应用于易感者-感染者-康复者-死亡者（SIRD）模型，以识别 2020年3月至2021年9月期间斯威士兰王国新冠肺炎的行为动力学。他们发现，即使训练数据量极少，其物理信息神经网络的表现仍优于所有其他数据分析模型 [6]。

Kharazmi 等人 [7] 研究了不同的整数阶、分数阶及时滞模型，这些模型均表示为常微分方程组。为了分析新冠肺炎在纽约市、罗得岛州、密歇根州以及意大利的过往动力学，他们使用了物理信息神经网络；据报道，该网络能够同时进行参数推断，并模拟已观测和未观测到的动力学。结果表明，纯统计方法通常不适用于流行病学动力学的长期预测；整数阶模型似乎比 Pang 等人最初提出的分数阶模型更稳健 [8]。此外，他们指出，没有任何模型能够准确捕捉长期大流行期间出现的所有动力学，但能够在训练过程中调整关键参数的模型或许能带来更有用的预测 [7]。

Cai、Karniadakis 和 Li 使用新颖的分数阶物理信息神经网络（fPINNs）深度学习框架，对易感者–暴露者–感染者–移除者（SEIR）模型中的未知模型参数进行了校准，以便对奥密克戎变体引发的新冠肺炎动态作出可靠的短期预测 [9]。所用数据来自中华人民共和国国家卫生健康委员会，涵盖 27th 2022年2月至2022年4月底。例如，预测能够捕捉新增感染病例趋势的突然变化。

另一方面，在一般的 PINN方法方面，近期多篇出版物都认识到 PINN训练具有多目标特性。Rohrhofer 等人 [10] 通过扫描权重区间，分析了加权和目标中数据损失与模型损失的不同权重所产生的影响。Jin 等人 [11, 第 4.4 节] 也在湍流通道流实验中研究了权重的影响，并通过手动调整权重来改善结果。最后，Wang 等人 [12, 算法 2。1]提出了一种称为“PINN 学习率退火”的自适应规则，用于在训练过程中在线选择权重。其基本思路是在模型训练期间利用反向传播梯度统计自动调整权重，从而适当平衡损失函数中的所有项。他们分别针对扩散方程和纳维–斯托克斯方程得到的数值结果，令人瞩目地展示了权重选择对训练成效的影响。事实上，多次训练运行均未能取得最优结果，因此仍有改进多目标训练方法的空间。当物理模型与数据高度吻合时（例如，数据由当前模型人工生成的情况），便可以寻求一种同时最小化数据损失和模型损失的“理想”解。

Maddu 等人 [13]提出了一种多目标下降法，采用逆狄利克雷策略自适应地更新权重，以避免过早终止。虽然他们没有讨论收敛保证，但其数值结果表明，与近期针对 PINN训练的多目标下降方法改进版本 [14, 15] 相比，该方法表现良好。随机多目标梯度下降算法最初用于一般的神经网络训练[17]。我们还提及自适应物理信息神经网络 [18]，以及将损失权重视为超参数的 PINN训练方法 [19]。

在更一般的设置下，文献 [20] 提出了多目标训练方法，用于在图像识别场景中权衡数据损失与正则化项。所考虑训练目标的不同特性（斜率和曲率）通过以下方式加以处理：在随机多梯度下降方法 [17] 中引入剪枝策略，并结合自适应加权和标量化—

与区间二分法相结合。后者有助于识别帕累托前沿上的有利膝点解。

本文组织如下：第2节介绍用于新冠肺炎预测的仓室模型。首先，第2.1节解释SIR模型，以帮助理解流行病建模的基本知识。随后，第2.2节定义仓室模型SVIHR模型的常微分方程组，并说明所采用的转移率和传播率。所有模型参数均列于表1。

第3节专门介绍方法方面的进展。第3.1节介绍非标准有限差分（NSFD）方法，解释该格式的概念，推导所谓的分母函数，并建立SVIHR模型的NSFD格式。第3.2节介绍物理信息神经网络（PINN），重点讨论损失函数和所提出的神经网络结构。第3.3节介绍一些双目标优化相关内容，以考察通过双目标PINN训练方法得到的帕累托前沿。最后，第3.4节介绍一种二分搜索方案，旨在快速找到近似理想的帕累托最优解，并为数据损失与残差损失之间更合适的权衡提供依据。

我们在第 4 节展示数值结果。在第 4.1 节中，我们使用德尔塔变异株疫情期间生成的数据作为训练数据，对感染数据进行短期预测，以预测第一波奥密克戎疫情。在第 4.2 节中，我们继续应用二分搜索方案，探讨其逼近帕累托前沿的能力，并确定合理的权衡解。最后，在第 4.3 节中，我们使用大部分可用数据作为训练数据进行长期预测，以预测德尔塔波。我们使用二分搜索，为两个目标函数寻找较低成本的加权参数。第 5 节给出结论，并展望未来工作。

2. 用于新冠肺炎预测的仓室模型

本文在 PINN训练期间用于计算残差损失的仓室模型是易感者-已接种者-感染者-住院者-康复者（SVIHR）模型，由 Treibert 和 Ehrhardt 在 [21] 中提出。为确保内容完整，我们在第 2.2 节再次简要推导该模型。SVIHR模型以 Kermack 和 McKendrick 在 1927 [22], 中提出的基本易感者-具有传染性-康复者 （SIR）模型为基础，扩展了 SIR 模型，纳入已接种者和住院者两个舱室。第 2.1 节对数理流行病学中的 SIR 模型作了简要介绍。

在比较数据驱动的PINN方法与近似SVIHR模型的非标准有限差分方法时，前者采用了独立的训练数据集；比较考察了两种方法对感染人数和住院人数的预测质量。

Treibert、Brunner和Ehrhardt [23] 重点关注非标准有限差分格式在易感者-已接种者-感染者-重症监护-死亡-康复者（SVICDR）模型中的表现。本文考察了修改参数边界对预测患病率的影响，同时纳入德国疫情数据、所研究时间窗口内呈指数增长的接种率，以及三角函数形式的接触率和隔离率函数。结果表明，非标准有限差分方法仅凭数学模型和所定义的参数即可预测全局峰值，而无需依赖此前观察到的传染病传播情况。

本文基于 [21]的SVIHR模型展开。我们提出了一种新颖的PINN方法，该方法基于来自 [24, 25, 26] 的更新数据，同时纳入短期数据和长期数据进行预测。改进的网络架构与基于动力学的参数估计相结合，后者融合了NSFD和PINN方法。所考虑应用的一个典型特点是，测量数据与模型预测之间往往存在显著偏差：虽然SVIHR模型能够捕捉新冠肺炎感染的动态，但无法反映数据中的波动，而这些波动可能由多种不同原因引起

。我们采用一种自适应方法，在训练过程中分析数据损失与残差损失之间的权衡。对PINN训练采用双目标视角，可半自动地针对具体问题确定最优加权参数。

由于目前我们的目标并非确定无症状个体在感染者总数中的比例，因此我们不设置单独的无症状感染者舱室，而是假设感染者至少会出现非常轻微的症状。可以通过调整模型中的传播率来调节感染者的传染性。我们的模型适用于不同的疫苗接种和传播情景。

2.1. 数理流行病学中的*SIR*模型

基本SIR模型由易感者（S）、感染者（I）和康复者（R）三个舱室组成。我们用 $K(t)$表示时间 t时舱室 $K \in \{S, I, R\}$ 的规模，其中一个时间单位为一周。易感者尚未感染，但可能患病。在基本SIR模型中，感染者可能会传染给易感者，即假定他们具有传染性（且不存在任何延迟），并且可能出现症状，也可能没有症状。康复者已战胜疾病，假定既不具有传染性，也没有患病。

时刻 t 的人口总数记为 $N(t)$。满足方程

$$N(t) = S(t) + I(t) + R(t) \quad \text{with } N\colon [0, T] \to \mathbb{N},$$

这意味着在每个所考虑的时点，系统中的个体数量等于各舱室人数之和 $t \in [0, T]$。系统 (2) 必须具有初始条件 $S_0 = S(0)$、 $I_0 = I(0)$、 $R_0 = R(0) = 0$ ，才能得到良好定义 [27,p. 11]。假设种群规模 $N(t)$ 恒定，即 $N(t) = N$，且 $N(t)$ 的导数为零，这意味着系统不考虑招募率 Λ 或自然死亡率。

令 p 表示与易感个体接触后导致传播的概率，并令 ζ 表示人均接触率，即一个具有传染性的个体所进行的接触次数。那么，ζN 表示该具有传染性的个体单位时间内进行的接触次数，而 $\zeta N \frac{S}{N}$ 表示一个具有传染性的个体单位时间内与易感个体发生的接触次数。此外，我们将传播率常数定义为 β [27, p. 10]，其表达式为

$$\beta = p\,\zeta\,. \tag{1}$$

关于传播率及其他相关参数（包括时变模型等）的更详细讨论，请参阅 [23]。

如果 $I(t)$ 表示时刻 t 的感染者人数（患病率），那么 $\beta S I$ 表示单位时间内新感染的人数（发病率）。如果 ω_I 是康复率，则可得到以下常微分方程组，用以描述 SIR模型 [27, 第 11] 页：

$$\begin{aligned}
\frac{\mathrm{d}S(t)}{\mathrm{d}t} &= -\beta\, I(t)\, S(t),\\
\frac{\mathrm{d}I(t)}{\mathrm{d}t} &= \beta\, I(t)\, S(t) - \omega_I\, I(t),\\
\frac{\mathrm{d}R(t)}{\mathrm{d}t} &= \omega_I\, I(t).
\end{aligned} \tag{2}$$

对于方程 (2) 中的模型，在所考虑的流行病中感染者人数的最大值不超过

$$I_{\max} = -\frac{\omega_I}{\beta} + \frac{\omega_I}{\beta}\,\ln\!\big(\frac{\omega_I}{\beta}\big) + S_0 + I_0 - \frac{\omega_I}{\beta}\,\ln\!\big(S_0\big)\,. \tag{3}$$

Let

$$F(t) = 1 - e^{-\omega_I t}, \quad t \geq 0 \tag{4}$$

表示在时间间隔 $[0, t)$ [27, p. 11]. 内康复／离开感染舱室的概率。函数 $F(t)$, 其中 $F(t) = 0$ for $t < 0$, 是一个概率分布。于是, $f(t) = \mathrm{d}F(t)\,\mathrm{d}t$ 即为相应的概率密度函数：

$$f(t) = \omega_I\, e^{-\omega_I t} \quad \text{for} \quad t > 0, \text{ and} \quad f(t) = 0 \quad \text{for} \quad t \leq 0. \tag{5}$$

如果 X 表示在感染舱室中停留的平均时间，那么在感染舱室中的平均停留时间可以用一阶矩计算

$$\mathbb{E}\big[X\big] = \int_{-\infty}^{\infty} t\, f(t)\, dt = \int_{0}^{\infty} t\, \omega_I\, e^{-\omega_I t}\, dt = \frac{1}{\omega_I}\,. \tag{6}$$

对于 SARS-CoV-2，具有传染性的平均时长尚无明确定义。若病情为轻度或中度，症状出现后的十天内，传染性会明显下降。传染性与检测结果呈阳性应予以区分：感染后数周仍可能检测呈阳性，但此时通常已几乎不具有传染性 [28]。

在这种基本形式的SIR模型中，假设人群是封闭的，即没有个体从外部进入或离开任何舱室；康复者则具有完全免疫力，因此永远不会再次感染[27,p. 13]。

*2.2. SVIHR*模型

在文献[21]中，SIR模型增设了一个住院人员舱室H 和一个已完成全程接种（即至少接种两剂）人员舱室 V。数据来自罗伯特·科赫研究所（RKI） [24, 25] 和德国新冠疫苗接种仪表板 [26], ，并据此计算出参数值和舱室规模，以下将其称为报告的舱室规模或报告数据。

感染者会持续感染 T_I 天，之后康复；从感染者中转出的所有个体中，有 ξ 的比例会住院。感染者中具有传染性的个体每日或每周的确切人数并不明确。本文使用罗伯特·科赫研究所（RKI）登记的感染人数，计算所考虑的所有日历周内舱室 I 的报告人数。该人数基于感染者中具有足够传染性、通常可通过快速抗原检测检出病毒的个体人数。具有传染性和不具有传染性的感染者在舱室 I 中合并计算。个体的总体

传染性程度取决于病毒的传播能力，并纳入传播率中。

根据罗伯特·科赫研究所的数据，传染性的具体持续时间尚未明确定义，但在最初症状出现前后，传染性最强；而从最初症状出现起，至多 10 天后，传染性会大幅下降（假设疾病进程为轻度或中度） [28]。我们设定$T_I = 1.2$ 周，即 8.4 天，以涵盖症状初次出现到接受检测之间 1 至 2 天的短暂间隔。参数 ω_1 表示单位时间（周）内个体从舱室 I 转入舱室 R的速率。其定义为

$$\omega_1 = \frac{1-\xi}{T_I}\,. \tag{7}$$

个体单位时间内进入舱室 H 的速率 η 定义为

$$\eta = \frac{\xi}{T_I}. \tag{8}$$

因此，我们假设每周有 $\omega_1 I\,(t)$ 人康复， $\eta I\,(t)$ 人住院。这里假设住院人员因处于隔离状态而不会感染易感个体。假设他们从住院时起仍处于感染状态 T_H 天。在一项针对 1,426 名患有急性呼吸道疾病的新冠肺炎患者的德国学术调查中，观察到平均住院时间为 10 天 [28]。因此，我们设定 $T_H = 1.5$ 周，即 10.5 天。

接种疫苗舱室 V包含所有已接种 COVID-19 疫苗的易感个体。个体以速率 $\mathcal{V}$从舱室 S 转入该舱室。如果疫苗接种不能确保完全免受感染，我们称之为渗漏型疫苗接种。由于假定疫苗存在渗漏，所有已接种个体感染的概率都低于舱室 S中的易感个体。如果假设使用全有或全无型疫苗，那么每单位时间内，易感者中有特定比例 t的人会完全免受感染，而其他易感个体则不会获得任何保护。令 κ 表示接种疫苗后的残余感染概率

。那么，已接种个体转入感染者舱室 I 的速率为 $\kappa\beta I(t)$。

此外，我们加入了一个恒定的系统流入，即所谓的招募率 Λ（例如可能被感染的新个体出生），以及自然死亡率 μ。由于本文将招募率和自然死亡率视为相等，因此二者均设为零 [21]，但仍将其纳入常微分方程组，以便在非标准有限差分格式中正确推导分母函数，见第 3.1 节。这样，总人口规模便保持恒定。相应的常微分方程组形式如下：

$$
\begin{aligned}
\frac{\mathrm{d}S(t)}{\mathrm{d}t} &= \Lambda - \beta I(t)S(t) - (\mathcal{V}+\mu)\,S(t),\\
\frac{\mathrm{d}V(t)}{\mathrm{d}t} &= \mathcal{V}\,S(t) - \kappa\beta I(t)S(t) - \mu\,V(t),\\
\frac{\mathrm{d}I(t)}{\mathrm{d}t} &= (1+\kappa)\beta I(t)S(t) - \big(\eta+\omega_1+\mu\big)\,I(t),\\
\frac{\mathrm{d}H(t)}{\mathrm{d}t} &= \eta\,I(t) - (\omega_2+\mu)\,H(t),\\
\frac{\mathrm{d}R(t)}{\mathrm{d}t} &= \omega_1\,I(t) + \omega_2\,H(t) - \mu\,R(t).
\end{aligned}
\tag{9}
$$

系统 (9) 在简单系统 (2) 的基础上，增加了描述舱室V（t）和 H（t）流入与流出的微分方程，分别表示流入和流出舱室V 或 H的过程，以及招募率 Λ 和自然死亡率 μ。请注意，(9) 中关于 V（t）的方程只能近似实际动态，因为接种疫苗人员即使感染后仍保持已接种状态。我们以易感个体舱室为参照，通过项 $\kappa\beta I$（t）S（t）来近似相应的变化率。采用非标准有限差分格式和物理信息神经网络方法进行的初步数值测试表明，使用 $\kappa\beta I$（t）V（t）不会显著影响预测质量。

容易证明，对于正参数和正的初始数据，方程 (9) 的解在所有时刻都保持为正。我们称之为“正性”。现在令 $\Lambda = \mu = 0$，并将 (9) 中的所有舱室相加，可得总人口是一个随时间守恒的量。稍后讨论非标准有限差分格式时，我们将再次讨论这一“守恒性质”。这种

设定是一种可接受的简化，因为人类出生和死亡的时间尺度远长于 C
OVID-19疫情波的时间尺度。

图1展示了式(9)所描述的动力系统。蓝色箭头从一个舱室指向另一个舱室，表示状态转移。表1列出了模型参数

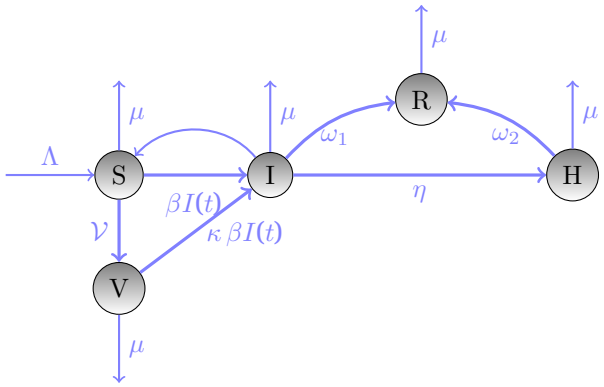


图1：SVIHR模型中各舱室及其相互关系示意图

定义和取值。需要指出的是，本文使用的SVIHR模型参数值（见表1）与文献[21]中的参数值不同。

3. 在PINN方法中寻找最优权重

我们实现了一种物理信息神经网络（PINN），该网络同时根据德国COVID-19数据和上述第2.2节介绍的SVHIR模型进行训练。我们通过针对报告数据进行误差计算来验证这一技术。第3.2节说明了物理信息神经网络的方法和结构。此外，我们将使用PINN生成的情景与使用非标准有限差分格式生成的情景进行比较。第3.1节简要介绍了非标准有限差分格式及其在SVIHR模型中的应用。

由于物理信息神经网络根据两个损失项进行训练，我们采用双目标方法来考察加权参数对各损失项的影响。我们将这两个损失视为相互独立的目标函数，并希望找到

Parameter	**Definition**	**Parameter Value used in 4.3**	**Parameter Value used in 4.1 and 4.2**
N	total population size	83,100,000 [29]	83,100,000 [29]
β	transmission risk parameter	*determined via NSFD*	*determined via NSFD*
κ	residual infection prob. after vaccination	*determined via NSFD*	*determined via NSFD*
T_I	average length of infection period	1.2 [28]	1.2 [28]
T_H	average length of hospitalization period	1.5 [28]	1.5 [28]
$\mathcal{V}$	vaccination coefficient	0.0159 [26]	0.0231 [26]
ξ	hospitalization coefficient	0.0862 [25]	0.0735 [25]
$\mathcal{M}$	mortality coefficient	0.0232 [24]	0.0142 [24]
ω_1	weekly recovery rate for infected people	0.7615	0.7721
ω_2	weekly recovery rate for hospitalized people	0.6512	0.6572
η	weekly hospitalization rate	0.0719	0.6125

Table 1: Parameters of the SVIHR model and their values in the implementation of the PINN.

加权参数，以近似帕累托前沿。因此，我们先在第 3.3 节介绍双准则优化的某些方面，再在第 3.4 节介绍一种二分搜索方案，以有效近似帕累托前沿。

3.1. 非标准有限差分格式

非标准有限差分格式最早见于 Mickens 于 1989 [30] 发表的一篇论文。其结构性质源于对某些特殊微分方程组的研究，对于这些方程组，不存在精确的有限差分格式。为求完整起见，这里再次介绍 SVIHR模型的非标准有限差分格式。类似的推导已在 [21] 中给出。

在非标准有限差分格式中，必须用恰当的离散类比来表示导数，即形如以下形式的非标准差商，参见 [31]

$$\frac{\mathrm{d}u(t)}{\mathrm{d}t} \to \frac{u_{n+1} - \psi(h)u_n}{\phi(h)}\,, \tag{10}$$

其中 $t_n = n\,h$, u_n 是 $u(t_n)$ 的近似，而 $\psi(h) = 1 + \mathcal{O}(h)$。是分母函数，下文将对此作更详细的解释。在非标准有限差分格式中，采用这一较为一般的时间离散（10），我们的目标是模拟解的渐近长期行为。如果一个一阶微分方程组的数值格式满足以下至少一个条件，则称其为非标准有限差分格式 [31]:

- 导数的离散表示通常必须具有非平凡的分母函数。在此，系统中的一阶导数通过广义前向差分法 $\frac{\mathrm{d}u_n}{\mathrm{d}t} \approx \frac{u_{n+1}-u_n}{\phi\,h}$ () 进行近似，其中 $u_n \approx u(t_n)$ 和 $\phi \equiv \phi(h) > 0$ 是所谓的分母函数，使得 $\phi(h) = h + \mathcal{O}h^2$，其中 h 为步长。

- 有限差商的相容阶应与微分方程中相应导数的阶数相同。

• 非线性项通过非局部离散表示进行近似，例如网格上若干点的适当函数，如$u^2(t_n) \approx u_n u_{n+1}$ 或 $u^3(t_n) \approx u_n^2 u_{n+1}$。

• 对于常微分方程和/或其解成立的特殊条件，也应对差分方程模型和/或其解成立，例如底层常微分方程组的平衡点。

与传统差分方法不同，非标准有限差分格式关注的不仅是稳定性和收敛阶，还包括定性性质，即离散模型（非标准有限差分格式）对底层连续模型最重要性质的再现程度。非标准有限差分格式保持正性，并满足 $\Lambda = \mu = 0$ 的守恒律，从而确保格式的稳定性。所提出的非标准有限差分格式中也包含常微分方程模型（9）的平衡点。

接下来，我们将为系统（9）的非标准有限差分离散化推导一个恰当的分母函数 $\phi(h)$。选择该函数的目的是使数值解呈现出与解析解相同的渐近行为。为此，我们考虑常微分方程组（9）的总人口 $N = S + V + I + H + R$。将方程（9）各式相加，可得到一个描述总人口 N 动态变化的微分方程：

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = \Lambda - \mu\, N(t)\,, \tag{11}$$

其求解方法为

$$N(t) = \frac{\Lambda}{\mu} + \left(N(0) - \frac{\Lambda}{\mu}\right) e^{-\mu t} = N(0) + \left(N(0) - \frac{\Lambda}{\mu}\right)(e^{-\mu t} - 1), \tag{12}$$

其中 $N(0) = S(0) + V(0) + I(0) + H(0) + R(0)$。与（11）类似，将（9）的非标准有限差分离散化方程相加可得

$$\frac{N^{n+1} - N^n}{\phi(h)} = \Lambda - \mu\, N^{n+1}, \tag{13}$$

i.e.

$$\begin{aligned} N^{n+1} &= \frac{N^n + \phi(h)\Lambda}{1+\phi(h)\,\mu} = N^n - \Big(N^n - \frac{\Lambda}{\mu}\Big)\frac{\phi(h)\,\mu}{1+\phi(h)\,\mu} \\ &= N^n + \Big(N^n - \frac{\Lambda}{\mu}\Big)\Big(\frac{1}{1+\phi(h)\,\mu} - 1\Big). \end{aligned} \tag{14}$$

通过比较方程（13）与方程（12）的离散化版本，即可推导出分母函数，也就是

$$N^{n+1} = N^n + \Big(N^n - \frac{\Lambda}{\mu}\Big)\,(e^{-\mu h} - 1), \quad h = \Delta t, \tag{15}$$

其中，（正）分母函数定义为

$$\frac{1}{1+\phi(h)\,\mu} = e^{-\mu t}, \tag{16}$$

i.e.

$$\phi(h) = \frac{e^{\mu h} - 1}{\mu} = \frac{1 + \mu h + \frac{1}{2}\mu^2 h^2 + \ldots - 1}{\mu} = h + \frac{\mu h^2}{2} + \ldots = h + \mathcal{O}(h^2). \tag{17}$$

现在，利用式（10）中的分母函数，可以建立 NSFD 离散化格式。该离散化格式的隐式形式见方程（18）。其中， $\phi(h)$ 由式（17）给出

$$\begin{aligned} \frac{S^{n+1} - S^n}{\phi(h)} &= \Lambda - \beta\, I^n\, S^{n+1} - (\mathcal{V} + \mu)\, S^{n+1}, \\ \frac{V^{n+1} - V^n}{\phi(h)} &= \mathcal{V}\, S^{n+1} - \beta\,\kappa\, I^n\, S^{n+1} - \mu\, V^{n+1}, \\ \frac{I^{n+1} - I^n}{\phi(h)} &= \beta\,(1+\kappa)\, I^{n+1}\, S^{n+1} - (\eta + \omega_1 + \mu)\, I^{n+1}, \\ \frac{H^{n+1} - H^n}{\phi(h)} &= \eta\, I^{n+1} - (\omega_2 + \mu)\, H^{n+1}, \\ \frac{R^{n+1} - R^n}{\phi(h)} &= \omega_1\, I^{n+1} + \omega_2\, H^{n+1} - \mu\, R^{n+1}\,. \end{aligned} \tag{18}$$

我们可以对该方案进行改写，以获得其显式变体，如式（19）所示。根据显式表示，我们可以推断该方案

保持正性。

$$
\begin{aligned}
S^{n+1} &= \frac{S^n + \phi(h)\Lambda}{1 + \phi(h)\,(\beta\, I^n + \mathcal{V} + \mu)},\\
V^{n+1} &= \frac{V^n + \phi(h)\, S^{n+1}\,(\mathcal{V} - \beta\,\kappa\, I^n)}{1 + \phi(h)\,\mu},\\
I^{n+1} &= \frac{I^n}{1 + \phi\,(\eta + \omega_1 + \mu - \beta\,(1+\kappa)\, S^{n+1})},\\
H^{n+1} &= \frac{\phi(h)\,\eta\, I^{n+1} + H^n}{1 + \phi\,(\omega_2 + \mu)},\\
R^{n+1} &= \frac{R^n + \phi(h)\,(\omega_1\, I^{n+1} + \omega_2\, H^{n+1})}{1 + \phi(h)\,\mu}\,.
\end{aligned}
\tag{19}
$$

计算必须严格按照这一顺序进行。在这类流行病模型中出现的所有参数始终都是非负的。

注 1. 观察式（18）的第三个方程，人们或许会疑惑为何不采用时间离散 $I^n S^{n+1}$，而采用 $I^{n+1}S^{n+1}$。这里有两种观点，各有道理，难分对错。采用这个“新”版本，在离散层面可以严格满足守恒性质。另一方面，采用“旧”版本 $I^{n+1}S^{n+1}$（隐式计算时，仅在 I^{n+1} 中使用；S^{n+1} 已在第一行中计算），则能保证正性守恒，因而也能保证稳定性性质，而这更为重要。因此，我们选择了第一种方案，并需要接受总人口守恒上的微小扰动。

关于一个实际应用，其中标准求解器会给出错误的负解，而这一问题可以通过非标准有限差分格式解决，请参阅 [32]。

3.2. 用于舱室模型的物理信息神经网络

物理信息神经网络 (PINN) 是一种将动力系统定律纳入深度学习框架的神经网络。对于由偏微分方程支配的高维问题，机器学习已成为数值离散的一种替代方法。然而，训练深度神经网络所需的充足数据未必总是可得。在这种情况下，可以通过纳入以下信息来弥补数据缺失：

通过施加动力系统的物理定律而获得的额外信息 [33]。这类定律可以用偏微分方程或常微分方程来描述。本文考虑的一个例子是，在流行病期间，人群在不同的感染或未感染状态之间转换，此时可以应用动力系统。

物理信息神经网络可以通过训练一个损失函数来近似微分方程的解，该损失函数纳入了初始条件和边界条件，以及所谓配点处的残差 [34]。物理信息神经网络不仅可以近似微分方程的解，还可以利用描述某一现实过程的微分方程组，以及描述该过程过去发展历程的时间序列数据集，从而预测其未来的发展。

相应神经网络的损失函数不仅包含所谓的数据误差 （与网络输出和所用报告数据之间的差异相关），还包含所谓的残差误差 （与常微分方程或偏微分方程相关）。

Olumoyin 等人 [35] 将一种前馈神经网络称为流行病学信息神经网络（*EINN*）；该网络通过将封锁等流行病学动力学因素纳入损失函数来发挥作用。EINN 扩展了适用于流行病学模型的物理信息神经网络，能够捕捉疾病传播的动力学及缓解措施的影响。损失函数还经过扩展，利用有关传染病的流行病学事实纳入时变率 [35]。

Shaier、Raissi 和 Seshaiyer [36] 将一种基于物理信息神经网络的神经网络称为疾病信息神经网络（*DINN*），该网络可应用于日益复杂的微分方程系统，用以描述各种已知的传染病。DINNs 可系统地应用于日益复杂的传染病微分方程控制系统。它们能够利用真实世界数据 [36] 有效学习疾病的动态，并预测疾病未来一个月的发展情况。

本文建立并应用的神经网络可以描述为

一种特殊类型的EINN或DINN，其基于一种特殊的流行病分舱模型，称为SVIHR模型，涉及人口中的易感者、已接种者、感染者、住院者和康复者。本文聚焦于德国人口，使用罗伯特·科赫研究所 [24, 25]和德国新冠疫苗接种仪表板 [26]提供的数据。（9）中包含的模型参数可以组成一个向量 ϑ:

$$\vartheta = [\beta, \mathcal{V}, \kappa, \xi, T_I, T_H, \mathcal{M}, T_H]^\top . \tag{20}$$

该向量分为固定参数 p_f（可直接根据观测数据估计）和可学习参数 p（对舱室变量的动态变化至关重要）：

$$\begin{aligned} p_f &:= [\mathcal{V}, \xi, T_I, T_H, \mathcal{M}, T_H]^\top, \\ p &:= [\beta, \kappa]^\top . \end{aligned} \tag{21}$$

我们利用 NSFD预测来改进对控制动态变化的参数的估计，即传播风险 β 和接种后的残余传播概率 κ。我们通过将 NSFD预测拟合到观测数据来估计这些参数的值，使预测峰值与观测峰值具有相同的量级。这种方法旨在兼顾两方面的优势：遵循物理模型的数值预测，以及高度依赖观测数据的物理信息神经网络预测。由于 NSFD预测的计算效率很高，因此可以轻松调整参数值，以适应不同的病毒变体和不同的疫情波次。

我们介绍适用于具有 n 个舱室的一般舱室模型的 PINN方法。在 SVIHR模型（9）中，我们有 $n = 5$。令

$$\hat{\mathcal{K}}(t) = [\hat{\mathcal{K}}^1(t), \ldots, \hat{\mathcal{K}}^n(t)]^\top$$

为 l 个时间点所报告舱室规模的归一化数据向量，即 $t \in \{t_1, \ldots, t_l\}$。在此处考虑的应用中，对于所有 $i = 1$、.....、$l-1$， $t_{i+1} - t_i$ 均为常数（对应一周）。令 T 表示所考虑时间窗口的上界，并令 $\mathcal{K}_p$: $[0, T] \to \mathbb{R}^n$ 表示

反映舱室规模 $\mathcal{K}_p(t)$ 随时间变化的向量值函数。于是，我们将式 (9) 右端项值组成的向量表示为

$$F(\mathcal{K}_p) = [F^1(\mathcal{K}_p), \ldots, F^n(\mathcal{K}_p)]^\top .$$

下标 p 表示该常微分方程组所依赖的可学习模型参数，参见 (21)，因此解也依赖于这些参数。于是，(9) 中的常微分方程组可离散化为

$$\frac{\mathrm{d}\mathcal{K}_p(t)}{\mathrm{d}t} - F(\mathcal{K}_p) = 0, \qquad t \in \{t_1, \ldots, t_l\} . \tag{22}$$

我们的物理信息神经网络

$$\mathcal{PINN}_p^W : [0, T] \to \mathbb{R}^n$$

用于通过训练 [37]期间的误差最小化，近似常微分方程组 (9) 的解 $\mathcal{K}_p$: $[0, 1] \to \mathbb{R}^n$ 。上标 W表示神经网络在前向传播与反向传播过程中使用的权重。在时间点 $t, t \in \{t_1, \ldots, t_l\}$，解表示为

$$\mathcal{K}_p(t) = [\mathcal{K}_p^1(t), \ldots, \mathcal{K}_p^n(t)]^\top ,$$

其中， $\mathcal{K}_p^j(t)$ 是物理信息神经网络在时间 t 下对第 j^{th} 个舱室的输出，并取决于方程 (21) 给出的参数 p 。参数 W 在神经网络的反向传播过程中进行优化，使得$\mathcal{PINN}_{\mathrm{WP}}$ 以最小二乘意义拟合报告数据 $\hat{}\mathcal{K}$ [37]。在第 i^{th} 个时间步 t_i, $i \in \{1, \ldots, l\}$，给定物理信息神经网络输出 $\mathcal{PINN}_{\mathrm{WP}}(t_i)$，我们计算通常定义的数据误差 ，如下所示

$$\mathrm{MSE}_{\mathcal{U}} = \mathrm{MSE}_{\mathcal{U}}(W) := \frac{1}{l} \sum_{i=1}^{l} \| \mathcal{PINN}_{\mathrm{P}}^{\mathrm{W}}(t_i) - \hat{\mathcal{K}}_p(t_i) \|^2 , \tag{23}$$

其中，我们在（23）中采用了 $\mathbb{R}^n$ 的欧几里得范数。

接下来，我们为物理信息神经网络的损失函数添加一项

$$\mathcal{F}_p(\mathcal{PINN}_{\mathrm{P}}^{\mathrm{W}}, t_i) := \frac{\mathrm{d}\,\mathcal{PINN}_{\mathrm{P}}^{\mathrm{W}}(t)}{\mathrm{d}t}\Big|_{t=t_i} - F_p\big(\mathcal{PINN}_{\mathrm{P}}^{\mathrm{W}}(t_i)\big) , \tag{24}$$

其中

$$\mathcal{F}_p(\mathcal{PINN}_p^W, t_i) = 0 \quad \text{for all } t \in \{t_1, \ldots, t_l\} \tag{25}$$

这意味着，通过施加额外约束，物理信息神经网络能够更可靠地求解给定的常微分方程组。神经网络输出的时间导数$\mathrm{d}\,\mathcal{PINN}^{\mathrm{W}}_{\mathrm{p}}(\quad t\quad)/\mathrm{d}t\big|_{t=t_i}$可以通过自动微分 [38]计算。于是，损失函数中物理信息部分的残差误差可表示为

$$\mathrm{MSE}_{\mathcal{F}} = \mathrm{MSE}_{\mathcal{F}}(W) := \frac{1}{l}\sum_{i=1}^{l} \|\mathcal{F}_p(\mathcal{PINN}^{\mathrm{W}}_{\mathrm{p}}, t_i)\|^2 . \tag{26}$$

我们引入一个超参数和权重因子 $\alpha \in [0,1]$ ，用于在损失函数中为数据损失和残差损失赋予权重。我们将总体损失函数定义为

$$\mathcal{L}_\alpha = \mathcal{L}_\alpha(W) := \alpha\ \mathrm{MSE}_{\mathcal{U}} + (1-\alpha)\ \mathrm{MSE}_{\mathcal{F}} \tag{27}$$

神经网络的最小化问题定义为

$$\min_W \big(\mathcal{L}_\alpha\big) . \tag{28}$$

需要简要指出，残差损失 $\mathrm{MSE}_{\mathcal{F}}$ 较小，并不一定意味着所得解接近式（9）的精确解。为此，还需要问题（9）是适定的，并且所用方案具有稳定性。所有实现均使用编程语言 *Python* 和深度学习框架 *PyTorch* 。我们采用了标准的网络架构，其中包含三个全连接隐藏层，每层有 30 个神经元，示意见图2。对于每个时间点 $t \in [0,T]$，网络都会返回向量 $\mathcal{K}_p(t) \in \mathbb{R}^5$中对应全部五个舱室规模的值。该网络架构借鉴了 Ben Moseley 的谐振子物理信息神经网络，见 [39], ，在网络复杂度方面达到了合理的折中。

训练中，我们使用了双曲正切激活函数和 Adam优化器。有关机器学习中下降方法的综述，请参阅 [40] 。我们还采用了学习率调度，在完成约 50 % 的训练迭代后大幅降低学习率。更具体地说，学习率

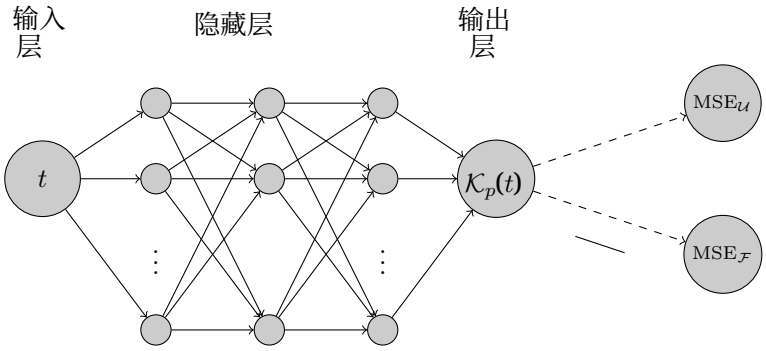


图2：所提出的物理信息神经网络的网络架构。每个隐藏层包含30个神经元。

$t(\kappa)$ 在第 $\kappa \in \{1, \ldots, \kappa_{\max}\}$ 次迭代时定义为

$$t(\kappa) = -(t_{\text{start}} - t_{\text{end}}) \frac{\exp\left(\frac{\kappa - 0.5 \cdot \kappa_{\max}}{0.08 \cdot \kappa_{\max}}\right)}{\exp\left(\frac{\kappa - 0.5 \cdot \kappa_{\max}}{0.08 \cdot \kappa_{\max}}\right) + 1} + t_{\text{start}}, \tag{29}$$

其中， t_{start} 和 t_{end} 分别表示初始和最终学习率。在我们的实验中，我们使用了 $t_{\text{start}} = 0.003$ 和 $t_{\text{end}} = 0.00015$。图3展示了经过 $\kappa_{\max} = 1000$ 次迭代的学习率调度。

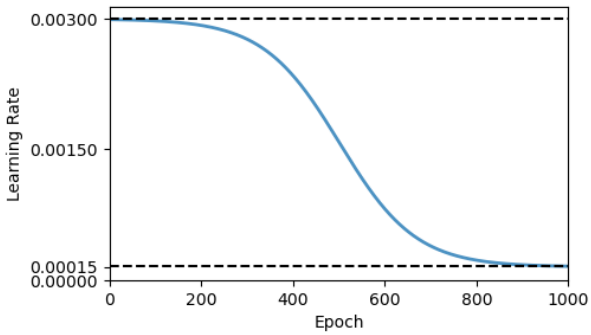


图3：我们采用的学习率调度，在1000个训练周期内逐步降低学习率。

数据损失是使用从罗伯特·科赫研究所 [24, 25]和德国疫苗接种仪表板 [26]获得的数据集计算的。计算数据损失均方误差$_{\mathcal{U}}$之前，先对这些数据进行预处理并重新缩放至单位区间。由于这会导致物理信息神经网络预测 $\mathcal{K}_p$相应地重新缩放，因此相应的

缩放因子在计算均方误差$_{\mathcal{F}}$时也被纳入常微分方程组（9）中。

数据涵盖了从2020年第10个日历周到2022年第14个日历周。根据给定的数据集，计算了每周病例住院率、病死率和疫苗接种率。罗伯特·科赫研究所将检测到SARS-CoV-2病原体的死亡个体登记为死于新冠肺炎的人。

加权损失函数 $\mathcal{L}_\alpha = \alpha$ 均方误差$_{\mathcal{U}} + (1-\alpha)$ 均方误差$_{\mathcal{F}}$由数据损失和残差损失项构成。如 [33], 所述，使用数据损失进行训练（即基于测量数据、非物理信息）被视为监督学习，而使用支配微分方程和残差损失进行训练（物理信息）则被视为无监督学习。

3.3. 相互冲突的训练目标：帕累托前沿与权衡分析

本文从双目标视角考察式（27）所述的优化问题。这为深入分析两个损失项均方误差$_{\mathcal{U}}$ 和均方误差$_{\mathcal{F}}$之间的权衡奠定了基础。我们不再考虑使用固定加权参数 $\alpha \in [0,1]$ 对这两个训练目标进行加权求和，而是分别考察这两个优化目标，并将它们纳入一个向量值目标函数 $\mathcal{L}$；该函数将每个可行解向量（W， p）映射为一个二维结果向量 $y = \mathcal{L}$（W， p） $\in \mathbb{R}^2$:

$$\min_W \mathcal{L}_\alpha(W) = \min_W \big(\mathrm{MSE}_{\mathcal{U}}(W), \mathrm{MSE}_{\mathcal{F}}(W)\big). \tag{30}$$

与前文一样， W 表示神经网络权重。为了分析双目标优化问题（30），我们首先回顾多目标优化领域的一些基本概念。关于该领域更详细的介绍，参见例如 [41]。

为此，我们将 Y 称为包含所有可能结果向量 $y = \mathcal{L}_\alpha(W) \in \mathbb{R}^2$的问题结果集(30)。解 $\hat{W}$ (即训练得到的一组 $\hat{}$ 神经网络权重 W)称为有效解或帕累托最优解，当不存在其他解 W (即不存在其他一组神经网络权重W)使得 MSE$_{\mathcal{U}}$(W) $\leq$MSE$_{\mathcal{U}}$($\hat{W}$) 且 MSE$_{\mathcal{F}}$(W) $\leq$MSE$_{\mathcal{F}}$($\hat{W}$)，其中

ˆ 两个不等式中至少有一个是严格的。如果 W 是帕累托最优解，则相应的结果向量 ˆ ˆ y =(MSE$_{\mathcal{U}}$(W)，MSE$_{\mathcal{F}}$(W))称为非支配点。因此，帕累托最优解是指无法在不使另一个损失函数恶化的情况下改善某一损失函数的解（即神经网络权重）。所有帕累托最优解（分别为非支配点）的集合记为 X_E （分别为 Y_N ）。注意， Y_N 也常被称为帕累托前沿。最终目标是高效地，即快速地逼近所谓的膝点解，这些解能使数据损失和残差损失都达到近似最优值。这通过专门针对神经网络训练设计的二分搜索策略来实现。

帕累托前沿的近似可得出膝点解的近似值。可以通过求解一系列参数化的单目标子问题（即所谓的 标量化）来找到它（另见例如 [41]）。为使这些子问题保持简单，我们采用加权和方法，从而得到目标函数为式（27）的单目标优化问题（28），即

$$\min_{W}\big(\mathcal{L}_\alpha(W)\big) = \min_{W}\big\{\alpha\ \mathrm{MSE}_{\mathcal{U}}(W) + (1-\alpha)\ \mathrm{MSE}_{\mathcal{F}}(W)\big\} \tag{28}$$

众所周知， [41] 对于所有加权参数 $\alpha \in (0,1)$，(28) 的最优解始终属于 Y_N。然而，反向命题只有在满足凸性假设时才成立。事实上，只要集合 Y 是$\mathbb{R}^2_{\geq}$-凸的，就可以通过改变加权参数 $\alpha \in (0,1)$生成完整集合 Y_N ；但在神经网络训练中，这一性质通常无法保证。我们回顾一下，若存在某个 $\alpha \in (0,1)$，使得 $W \in X_E$ 是 (28) 的最优解，则称帕累托最优解 $W \in X_E$ 为受支持的。所有受支持的有效解和受支持的非支配点分别记为 X_{sE} 和 Y_{sN}。请注意，所有受支持的非支配点都位于 Y 的凸包边界上（例如参见 [42]）。图4展示了一个示例：集合 Y中位于 $\mathbb{R}^2$ 内的受支持非支配点集合 Y_{sN}，以及一个非受支持的非支配点。选择 (28) 中的加权参数 α 既至关重要，又极具挑战性。若 α 取值过大，就会过分强调数据损失

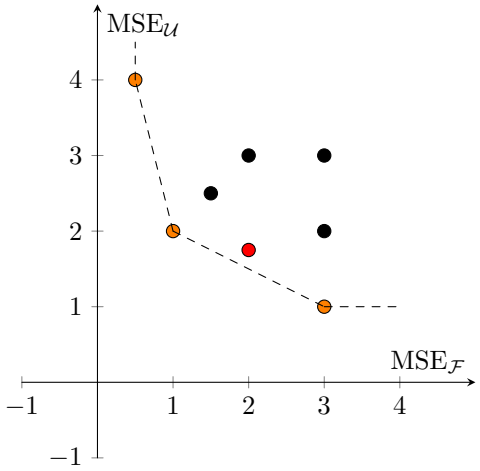


图4：神经网络训练得到的一组可能结果向量（均方误差$_{\mathcal{F}}$，均方误差$_{\mathcal{U}}$）的示意图，属于 Y_{sN} 的点以橙色表示，非受支持的非支配点以红色表示。虚线勾勒出结果向量集合 $Y + \mathbb{R}^2_{\geq}$的凸包。

因此往往会导致过拟合；若 α 取值过小，则可能过度优先考虑可能仅为近似的物理模型。尝试通过遍历潜在权重 $\alpha \in (0,1)$来近似帕累托前沿时，即使对于双目标凸问题，预先设定的加权参数也可能导致帕累托前沿上的点分布极不均匀，例如参见 [43]。此外，这种方法难以扩展到包含两个以上训练目标的高维问题。事实上，在搜索所有合理权重时，所需的加权向量数量 α （因而，针对加权和目标 (28) 进行的单独神经网络训练次数）会随着目标函数数量的增加而呈指数增长。

下一节中，我们将介绍一种有效的方法，能够自适应地选择加权参数，以确定具有两个优化目标时的膝点解（在我们的案例中，这两个目标是数据损失和残差损失）。我们强调，该方法可以推广到两个以上的目标函数（例如，用于偏微分方程的物理信息神经网络求解），因此有望显著减少确定膝点解所需的单次神经网络训练次数

近似理想的神经网络权重。

3.4. 用于自适应帕累托前沿近似的二分搜索

可以通过调整文献 [42]中描述的基于标量化的二分搜索算法，找出接近理想的神经网络解。其思路是自适应地计算加权参数 α，从而在最有希望的区域细化当前的帕累托前沿近似。该方法旨在根据问题的曲率和尺度进行自动调整，以便快速找到一组多样化的解，并迅速确定接近理想的膝点解。此外，二分搜索很容易融入交互式流程，使人们能够放大查看帕累托前沿中决策者最感兴趣的特定部分，参见例如 [44]。

以下思想可见于 [42]。在双目标情形下，二分搜索归结为求解一系列带有 $\alpha \in (0,\ 1)$ 的加权和标量化 (28)，并利用这样一个事实：在二维情形下，对于两个非支配点 y^r 和 y^s，有 $y_1^r < y_1^s$ 意味着$y_2^r > y_2^s$。随后，求解带有 $\alpha = (y_2^r - y_2^s)/c > 0$ 和$1-\alpha = (y_1^s - y_1^r)/c > 0$的加权和标量化 (28)，其中 $c = y_2^r - y_2^s + y_1^s - y_1^r$，以寻找位于 y^r 和 y^s之间的新受支持点。因此，加权参数 α 定义了连接 y^r 和 y^s 的线段的法向量，因为

$$\begin{aligned}\left(y^r - y^s\right)^\top \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix} &= \begin{pmatrix} y_1^r - y_1^s, & y_2^r - y_2^s \end{pmatrix} \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix} \\ &= c \cdot \begin{pmatrix} -(1-\alpha), & \alpha \end{pmatrix} \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix} = 0.\end{aligned}$$

图5左侧对此进行了说明。使用新的加权参数 α 求解加权和问题，会得到一个非支配点 y^t （如果该问题求得全局最优解），对此可能出现两种情况：

1. 如果 $(\alpha, 1-\alpha)^\top y^t < (\alpha, 1-\alpha)^\top y^r$，那么 y^t是一个新的受支持非支配点。由此生成两个新的子问题，其中一个由 y^r 定义

另一个由 y^t 定义，而 y^t 和 y^s。这种情况如图5右侧所示。

2. 如果 $(\alpha, 1-\alpha)^\top y^t = (\alpha, 1-\alpha)^\top y^r = (\alpha, 1-\alpha)^\top y^s$，那么 y^t 位于连接 y^r 和 y^s 的线段上，可以停止在此区间内搜索。

二分搜索按级推进，第1级包含由两个初始权重 α_1 和 α_2得到的加权和标量化的结果向量$\alpha_1 < \alpha_2$，以及一次二分步骤（相关加权参数的示意图见图5左侧）。在第2级中，对由当前帕累托前沿近似的凸包各线段所定义的所有权重进行加权和标量化（示意图见图5右侧）。搜索会持续进行，直到评估了预先设定数量的级，或者不再生成新的子问题。

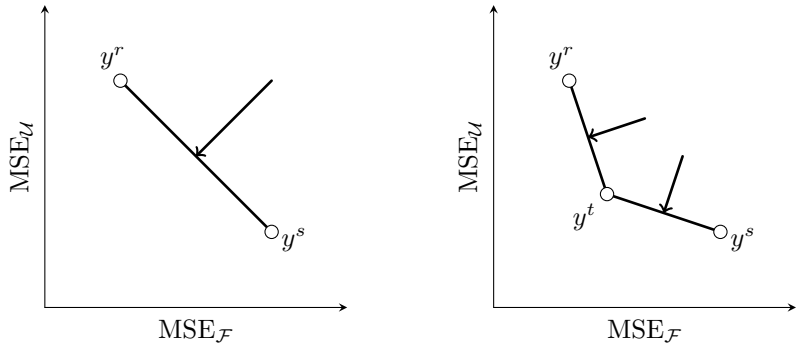


图5：左图：连接 y^r 和 y^s 的线段垂直于向量 $(\alpha, 1-\alpha)^\top$；此处将该向量乘以一个负标量，以展示最小化方向。右图：由于 $(\alpha, 1-\alpha)^\top y^t < (\alpha, 1-\alpha)^\top y^r$，新点 y^t 是非支配的，因此会生成两个新的子问题，见 [42,第 7] 页（原创示意图）。

算法 1 概述了二分增强二分搜索（BEDS）的实现，该方法通过偶尔执行二分步骤来增强二分搜索方案。它基于 [20] ，并考虑到这样一个事实：使用特定权重参数 α 进行神经网络训练可能会得到局部最小值，即训练（此处使用Adam优化器）可能会终止于局部最小值，而非一般二分搜索过程中假定的全局最小值。

算法 1： 二分增强二分搜索（BEDS）

数据： 训练数据、Adam 求解器的超参数设置、搜索深度（层级）、初始加权参数

α_1，$\alpha_2 \in (0, 1)$，$\alpha_1 < \alpha_2$，以近似极端解分别聚焦于均方误差$_{\mathcal{U}}$ 和均方误差$_{\mathcal{F}}$（见（27））

结果： 帕累托前沿及相应物理信息神经网络的近似参数

$\Lambda \leftarrow \{\alpha_1, \alpha_2\}$;

候选者 $\leftarrow \emptyset$;

对于 $l = 1, \ldots,$ 层级 **进行**

  对于 $\alpha \in \Lambda$ **执行**

    使用加权和目标 $\mathcal{L}_\alpha$进行训练;

    将目标向量（均方误差$_{\mathcal{U}}$，均方误差$_{\mathcal{F}}$）添加到 cand;

  删除 cand 中所有被支配点;

  按第二个目标函数（按递增顺序）对 cand 进行排序;

  如果 $l <$层级 **则**

    对于 $i \in \{2, \ldots, |候选|\}$ **执行**

      差异 $\leftarrow$ 候选(i) $-$ 候选$(i-1)$;

      $\alpha_{\text{new}} \leftarrow \left\{ \frac{\text{-差异}_1 \text{差异}_2 - \text{差}}{\text{异}_1} \right\}$;

      如果 $(0.999 - \alpha_{new}) < 0.001$ **或** $\alpha_{\text{new}} < 0.8$ **则** // 失败权重设置

        $\alpha_{\text{new}} \leftarrow (\text{succ}(\alpha) + \text{pred}(\alpha))/2$;

        // 用二分父区间替换未成功的 α 区间;

      $\Lambda \leftarrow \Lambda \cup \alpha_{\text{new}}$;

    按递增顺序排序 Λ

返回 cand 中的非支配点，以展示权衡;

为克服由此产生的数值困难，二分方案通过二分步骤得到增强，这些步骤会生成新的且有前景的——

需要时使用加权参数 α 。当新得到的权重落在原始搜索区间 $[\alpha_1, \alpha_2]$之外时，就会执行二分步骤。如果二分搜索计算出的斜率过小，就可能出现这种情况，从而导致数值问题。算法 1 会执行多次完整的训练运行，并在第 6 行之后重置网络的权重。变量 cand 初始化为空集，用于存储训练过程中针对损失函数 $\mathcal{L}_\alpha$ 的不同加权参数 α 生成的所有目标向量 $y =$(均方误差$_\mathcal{U}$、均方误差$_\mathcal{F}$)。

4. 结果

本节分为三个部分。在第 4.1 节中，我们对物理信息神经网络进行了验证。首先，我们使用德尔塔变异株时期（85th 至100^{th} 所考察的周，即 2021年第41周至2022年第4周）产生的数据，预测首轮奥密克戎变异株疫情（100th 至 104th 所考察的周，即2022年第4周至第8周）。我们使用非标准有限差分格式来确定 β 和 κ的拟合参数值。

在第 4.2 节中，我们展示二分搜索的结果，以探讨加权参数 α 对训练目标 (27) 的影响。我们讨论二分搜索近似帕累托前沿的能力，这有助于找到合理的权衡解。

最后，在第 4.3 节中，我们展示了一个使用长期训练数据集的结果，该数据集始于 2020 年 1 月大流行暴发之初。我们讨论了使用两个训练数据集得到的结果之间的差异。因此，我们区分了覆盖德国疫情暴发以来这段时期的训练数据集（称为长期训练数据），以及仅包含大流行期间某一特定波次甚至某个峰值数据的训练数据集（称为 短期训练数据）。

4.1. 使用短期训练数据生成的情景的验证

在本部分中，我们使用短期训练数据，并据此预测随后几周的感染者舱室 I 。我们使用短期数据的原因是，我们的物理模型旨在描述恰好出现一个峰值的一波

感染。由于 COVID-19 大流行期间出现了多波由不同病毒变体引起、且感染参数各异的疫情，因此仅使用一波数据进行训练来预测下一波，是将我们的模型纳入分析时更现实的做法。不过，我们将在第 4.3 节中展示使用更长训练期数据所得的结果，以作比较。值得注意的是，要预测感染率因奥密克戎传播而进一步上升，训练数据中必须包含涵盖这一感染率上升过程的数据。物理信息神经网络会根据底层数据集调整其训练参数。

我们首先使用非标准有限差分格式确定感染率 β 和 κ 的合适参数值。图6展示了我们用于训练和验证物理信息神经网络的时间范围内（归一化的）感染数据，以及使用非标准有限差分格式（见式（19））和表1右栏中的参数，即 $\beta = 0.00000001476$ 和 $\kappa = 0.001$，得到的预测。

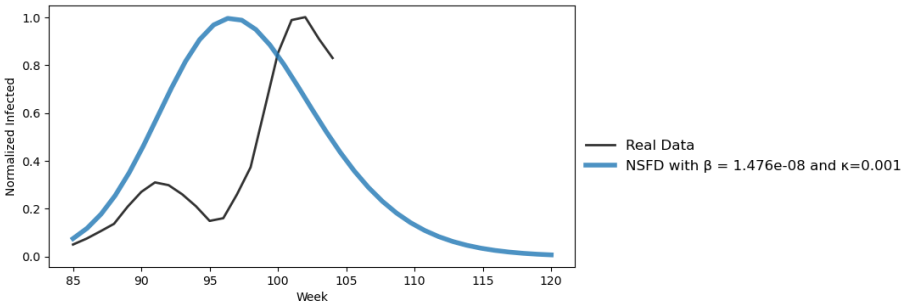


图6：从第 85th 个考察周到第 104th 个考察周的归一化感染人数，以及非标准有限差分格式得到的归一化感染部分。

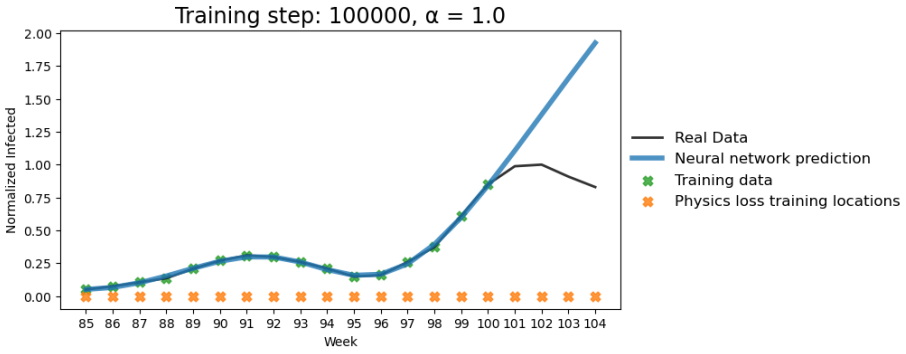
我们发现，这组参数既能很好地再现德尔塔波初期的斜率，也能再现奥密克戎变异株疫情的峰值。采用恒定传播风险的非标准有限差分格式显然只能预测一轮疫情中的一个峰值。我们注意到，在 NSFD 方法中加入随时间变化的传播率，有助于预测一轮疫情中的多个峰值，如 [21] 所示。为简单起见，本研究使用的物理信息神经网络或 NSFD 格式均未加入随时间变化的传播率；这将留待未来研究。

图 7 展示了短期预测的结果：物理信息神经网络使用从所考察的第 85th 周至第 100th 周的数据进行训练，并用于预测随后的四周（第 $101-104$ 周）。我们采用了三个不同的加权参数值 $\alpha \in \{1.0, 0.995, 0.99\}$，见公式（27）。每次训练均进行了 100000 次迭代。

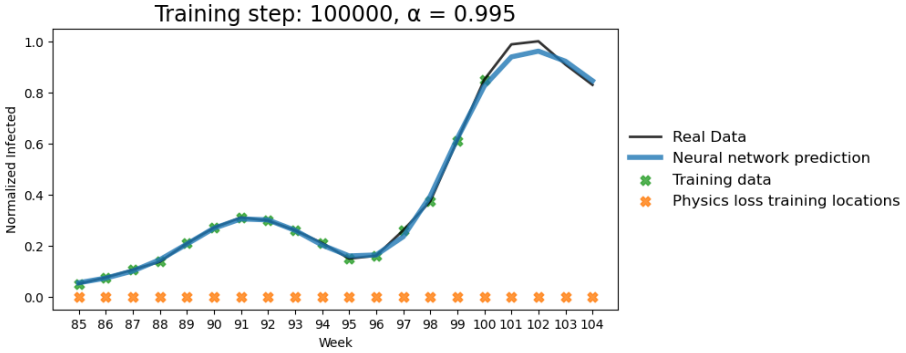
我们注意到，加权参数 α 对预测质量起着重要作用。在图 7(a) 中，$\alpha = 1$ 被选中，因此损失仅考虑均方误差 $_{\mathcal{U}}$。在这种情况下，物理信息神经网络几乎完美地拟合了训练数据，但由于缺乏有关系统物理特性的信息，导致过拟合，因此无法预测随后几周感染人数的下降。图 7(b) 显示，使用 $\alpha = 0.995$ 虽然对训练数据的拟合稍逊一筹，但能很好地预测随后四周的感染人数。最后，图 7(c) 显示，再次对均方误差 $_{\mathcal{F}}$ 赋予过高权重会导致预测效果不佳。这可以解释为：物理信息神经网络学到了德尔塔波的物理参数，而这些参数与奥密克戎变异株疫情的物理参数不同。因此可以看出，在物理模型只能部分解释真实动态的应用中，必须对数据损失赋予远高于残差损失的权重，才能获得良好的预测。

然而，完全忽略残差损失仍无法使我们的物理信息神经网络达到最佳性能。尽管没有任何数学模型能够最优地描述一种传染病，因为并非所有与传播相关的流行病学细节都已为人所知，但我们可以利用残差损失，将有关该疾病传播及其传播动力学的系统性知识纳入神经网络。由于 $\alpha = 1$ 在我们的验证试验中并未产生最小误差，因此纳入残差误差仍然是合理且有依据的。总的来说，受未知突变变异株、不断变化的干预措施或人群依从性，以及新的疫苗接种策略等因素影响，对未来疫情动态的预测存在诸多不确定性。

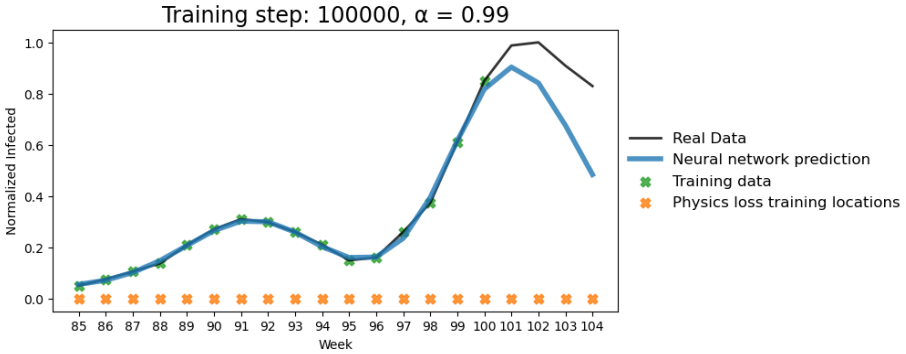
表 2 显示了我们的物理信息神经网络预测$\mathcal{K}_p^3(t_i)$中感染者舱室与报告感染数据 $\hat{}\mathcal{K}_p^3(t_i)$之间的均方误差



(a) 对训练数据完美拟合，但预测效果较差。



(b) 对训练数据拟合良好，预测效果也好。



(c) 如果均方误差$_{\mathcal{F}}$ 的权重过高，预测效果会变差。

图 7：从 85th 个纳入考虑的周到第104^{th} 周，根据报告数据（黑色）或采用损失项 MSE$_{\mathcal{U}}$ 和 MSE$_{\mathcal{F}}$ （蓝色）训练物理信息神经网络得到的感染人数，展示了加权参数 α的三种不同取值。训练数据集涵盖 2021年日历周第 41 周至 2022年日历周第 4 周（从85^{th} 到第 100th 个纳入考虑的周）。

α**的值**	1.0	0.999	0.998	0.997	0.996	0.995	0.994
$\mathrm{MSE}_{\mathrm{val}}$	0.0961	0.0553	0.0505	0.0031	0.0012	0.0003	0.0024
α**的值**	0.993	0.992	0.991	0.99	0.95	0.9	0.8
$\mathrm{MSE}_{\mathrm{val}}$	0.0016	0.0097	0.0109	0.0104	0.0772	0.0853	0.1030

表 2：根据加权参数 α，在考虑从 85th 到 104th 周的情况下，报告的感染数据与物理信息神经网络结果之间的均方误差。这可视为一次验证，因为除了用于训练的数据（第 85 至 100 周）外，还包括对第 100 至 104 周的预测误差。

考虑第 85 至 104 周（即整个有数据可用的时间范围，而不仅仅是用于训练的时间范围），均方误差按以下方式计算

$$\mathrm{MSE}_{\mathrm{val}} := \frac{1}{l}\sum_{i=1}^{l} \|\mathcal{K}_p^3(t_i) - \hat{\mathcal{K}}_p^3(t_i)\|^2 . \tag{31}$$

为了创建表 2，我们手动调整了加权参数 $\alpha \in$(0.8,1)。值得注意的是，最小误差是在 $\alpha = 0.995$（在所考虑的参数值中）时取得的。使用 $\alpha = 0.996$时，误差会增至四倍；使用 $\alpha = 0.994$ 时，误差则会增至八倍。不过请注意，各次训练运行不一定会以（全局）最优解结束，因此报告的误差值只能近似反映相应加权参数选择下可能达到的最佳误差。尽管如此，我们仍可观察到，在设置为 $\alpha = 0.9$ 或 $\alpha = 1.0$时，网络性能明显下降。例如，当加权参数 α 从 $\alpha = 0.99$降至 $\alpha = 0.9$时，均方误差$_{\mathrm{val}}$ 约增至八倍。

上述讨论表明，权重参数 α的选择对预测质量起着重要作用。需要注意的是，这里的预测质量是通过与真实数据进行比较来衡量的。因此，当加权和训练目标 (27) 中的数据损失均方误差$_{\mathcal{U}}$获得较高权重时，预测质量更高也就不足为奇了。然而，结果还表明，残差损失均方误差$_{\mathcal{F}}$ 不应被忽视。这

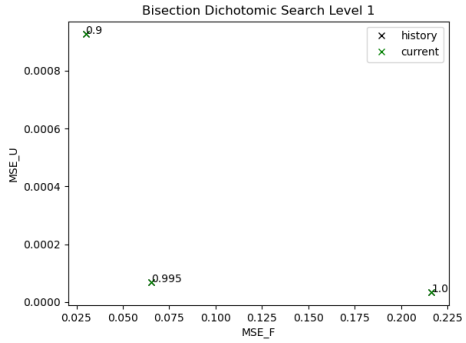
促使我们在下一节中使用二分搜索，对数据损失与残差损失之间的权衡进行更深入的分析。

4.2. 二分搜索的结果

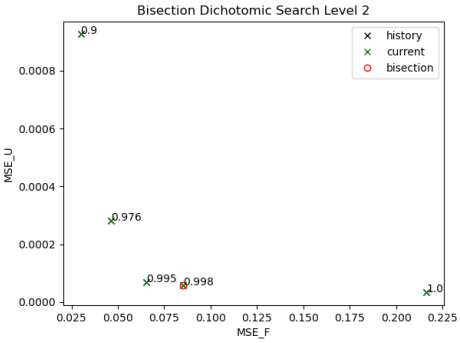
本节介绍使用算法 1 进行二分搜索的结果。根据前几节的讨论，我们重点关注从 2021 年第 41 个日历周到 2022 年第 4 周的短期训练数据（85th 至 100th 所考虑的周）。比较结果见第 4.1 节和图 7。图 8 展示了四个层级的帕累托前沿近似（分别基于 3、5、9 和 13 次训练运行）。每次训练运行均使用Adam优化器，训练 100000 个训练周期，并采用公式 (29) 中定义的学习率调度，其中 $t_{\text{start}} = 0.003$ 和 $t_{\text{end}} = 0.00015$。搜索从 $\alpha_1 = 0.9$和 $\alpha_2 = 1.0$开始。根据偏好或是否有额外信息，也可以使用其他搜索窗口。此外，搜索窗口还可用于放大帕累托前沿的特定部分。

结果证实，仅经过少量训练运行后，便近似得到了一个明显的帕累托前沿，其中包含多样的结果向量和清晰的膝点解。事实上，基于 5 次训练运行的第2级（见图 8(b)）已经能够对帕累托前沿进行粗略近似。在第3级（图 8(c)）中，经过 9 次训练运行后，我们已经得到了对帕累托前沿非常好的近似。最终的第4级基于 13 次训练运行（图 8(d)）。在配备 AMD Ryzen 7 3700X 8 核处理器和 NVIDIA GeForce RTX 3060 显卡的计算机上，每次训练运行耗时约 3 分钟，因此最终级别所需的计算时间约为 39 分钟。当然，每次训练运行的计算时间取决于具体问题和网络；然而，这里的目标是展示，仅需相对较少的训练运行次数，便可以深入了解该问题的完整帕累托前沿。

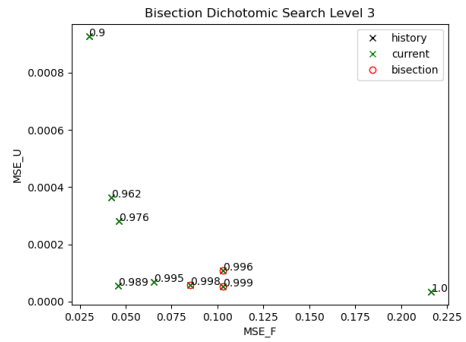
这表明，二分搜索方案仅需极少的训练运行，便成功获得了一组多样化的解，使决策者能够深入了解改变加权参数α 对各损失函数重要性的影响。



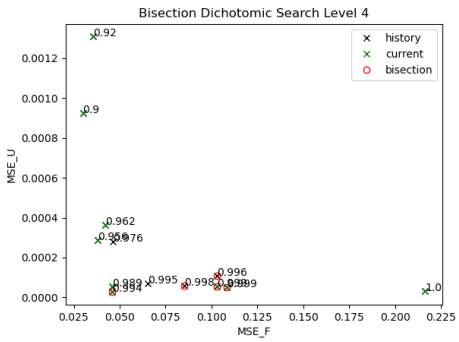
(a) 第1级：字典序优化和一次加权和（3 次训练运行）



(b) 第2级： 5 次训练运行, 二分法 used



(c) 第3级：9 次训练运行



(d) 第4级：最终级别，共进行 13 次训练运行

图 8：使用二分搜索进行帕累托前沿近似，初始学习率为 0.003，学习率调度的初始权重为 $\alpha_1 = 0.9$ 和 $\alpha_2 = 0.999$。从第3级开始即可识别出有利的权衡。

结果还表明，在不显著损失损失函数中均方误差$_{\mathcal{F}}$ 部分表现的前提下，可以大幅改善均方误差$_{\mathcal{U}}$ 。使用 $\alpha = 1$ 时，我们得到均方误差$_{\mathcal{F}} = 0.2176$ 和均方误差$_{\mathcal{U}} = 7.495 \cdot 10^{-5}$；而 α =0。994时，得到均方误差$_{\mathcal{F}} = 0.0471$ 和均方误差$_{\mathcal{U}} = 7.599 \cdot 10^{-5}$。因此，我们可以使用BEDS算法，快速、自动地确定物理信息神经网络损失函数中参数权重的有利权衡。

总体而言，可以看出， α 的变化会显著影响数据损失和残差损失的相对重要性。

4.3. 使用长期训练数据生成的情景验证

最后，我们希望利用现有的大部分数据进行长期预测。我们将所考察的第 1st 至第 89th 周（2020年第10周至2021年第45周）作为训练数据，并让物理信息神经网络预测随后出现的德尔塔波（所考察的第89^{th} 至第 95th 周，即2021年第45周至第51周）。首先，我们再次应用非标准有限差分格式，重新评估 β 和 κ的参数值。图9展示了这一过程的结果。

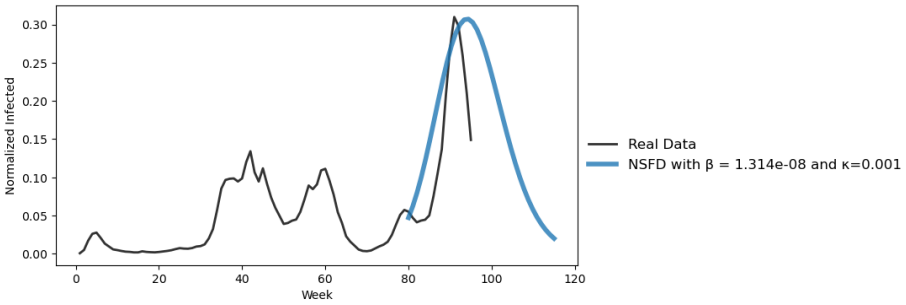
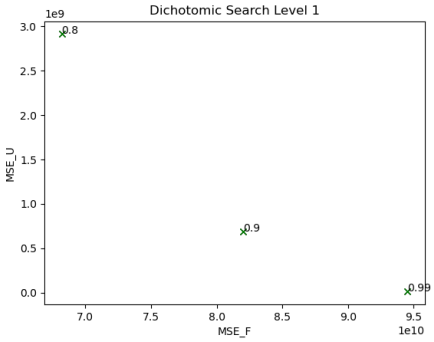


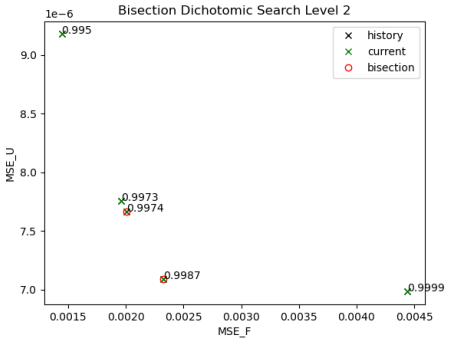
图9：从第 1st 个考察周到第 95th 个考察周的归一化感染人数，以及非标准有限差分格式的归一化感染部分。

我们选择 $\beta = 0.00000001314$ 和 $\kappa = 0.001$。使用这些值，我们开始另一轮二分搜索，以找到具有良好权衡效果的损失权重。由于此案例中的底层物理系统对数据的拟合不如前一次短期预测（长期数据包含多个感染峰值各不相同的波次），因此我们选择了一个更接近 $\alpha = 1$的搜索区间

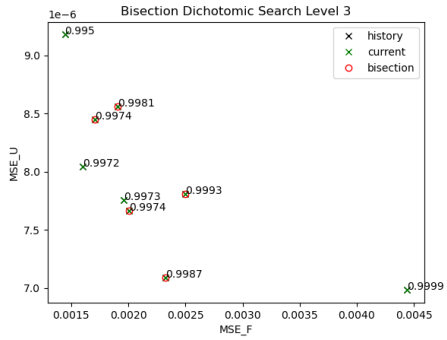
，以使用 $\alpha = [0.995, 0.9999]$ 对数据损失赋予更高权重。结果见图10。



(a) 第1级：字典序优化和一次加权和（3次训练运行）



(b) 第2级：5次训练运行，使用二分法



(c) 第3级：9次训练运行

图10：使用二分搜索近似帕累托前沿，初始学习率为0.003，并采用学习率调度，初始权重为 $\alpha_1 = 0.995$和 $\alpha_2 = 0.9999$。从第3级开始可以找到有利的权衡方案。

我们认为权重 $\alpha = 0.9987$ 是一个良好的权衡方案，它在均方误差$_{\mathcal{U}}$损失上没有太大损失，而且在均方误差$_{\mathcal{F}}$损失上远优于 $\alpha = 0.9999$ 。我们决定将这一权重用于长期预测。结果见图11。

我们注意到，与短期预测相比，这一预测的质量较低。鉴于底层的 SVIHR 系

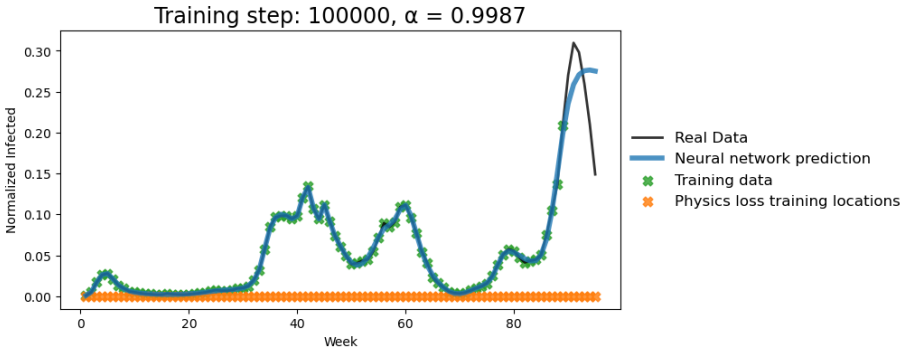


图11：由报告数据（黑色）或采用损失项 $\mathrm{MSE}_{\mathcal{U}}$ 和 $\mathrm{MSE}_{\mathcal{F}}$ （蓝色）训练物理信息神经网络得到的感染人数，时间范围为从所考虑的第 1st 周到第 95^{th}周，针对 $\alpha = 0.9987$。训练数据集涵盖 2020 年第 10 周至 2021 年第 45 周（从1^{st} 到所考虑的第 89th 周）。

统旨在模拟一次感染浪潮，这并不令人意外。为改善这些结果，需要纳入随时间变化的疫苗接种函数和传播率函数，这也是我们计划在未来工作中探讨的方向。

5. 结论与展望

我们研究物理信息神经网络（PINN），这是一种结合数据与物理知识的深度学习技术。本文以 COVID-19 感染率预测为案例研究。我们的主要贡献是为 PINN 训练中数据损失与残差损失之间的权衡提供了全新视角。我们提出了一种内在的双目标方法，能够有效识别近似理想的膝点解。此外，我们还采用逆向建模方法，根据相关微分方程组的数值解确定支配系统动态的模型参数。

在我们的方法中，数据损失是根据德国 COVID-19感染率数据计算的。物理信息神经网络（PINN）预测既有的易感者-已接种者-感染者-住院者-康复者（SVIHR）模型中各舱室的人数规模，重点关注感染者舱室。残差损失则

根据基于所提出的 SVIHR 模型的常微分方程组得出，该方程组以数学方式描述了受 COVID-19大流行影响的人群中不同舱室之间的转变动态以及传染病传播。我们提出了一种专为求解 SVIHR 模型数值解而设计的非标准有限差分格式。该格式的数值解用于估算传播风险 β 和接种后的残余传播概率κ，这两者共同支配着动态变化。

我们的结果表明，物理信息神经网络的预测质量高度依赖于所采用的数据，以及将两个相互冲突的损失函数通过加权和 $\mathcal{L}_\alpha$结合起来的加权参数 α。结合短期数据并优化加权参数 α，能够很好地预测首波奥密克戎变异株疫情。

该方法基于二分搜索，通过迭代逼近帕累托前沿，并以相对较少的训练运行识别接近理想的膝点解（即训练后的网络）。我们发现，在大多数情况下，总损失函数中加权参数 α 的首选值大于0.99，因此数据损失的权重高于残差损失。鉴于数据损失和残差损失的尺度不同，解读这一数值时需要谨慎。我们强调，无论损失项数量或应用领域如何，PINN训练的双目标方法都可以扩展应用于物理信息神经网络。理论上，可以通过扫描一系列合理的候选权重值来获得适当的加权参数 α；但这需要进行数量庞大的单独神经网络训练，且训练次数会随着损失函数数量的增加而呈指数级增长。

最后，我们采用 NSFD 参数估计和二分搜索，对 COVID-19 感染率进行长期预测，并使用大部分可用数据作为德尔塔波预测的训练数据。我们发现，即使是长期预测，PINN 也能给出合理的结果。然而，由于长期动态受到许多不同因素和措施的影响，此时物理模型的准确性更低，这也解释了为何其表现略逊于短期预测。

未来的工作将纳入随时间变化的疫苗接种函数和传播率函数，即 $\beta = \beta(t)$、$\kappa = \kappa(t)$，以考虑季节性和变异株相关的波动，参见 [45]。尤其是在长期预测中，应考虑传播率随时间的变化；此类预测的训练数据集涵盖数月，因此包含不同的突变、封锁措施和季节。为此，还必须考虑康复者舱室 R 到易感者舱室 S的回流。值得注意的是，我们在此仅处理了两个损失项，因此采用了双目标优化模型。多目标 PINN 训练的概念很容易扩展到考虑两个以上的目标函数。在更复杂的应用中，例如使用 PINN 求解偏微分方程（PDE），就会出现这种情况；例如，采用梯度增强物理信息神经网络（gPINN） [46], 时，损失项的数量会大于两个，因此需要采用多目标方法。例如，在 [46, 第 2.2], 节中，损失项的数量为 $3+d$，其中d 为空间域的维数。此时，必须识别相似且相互冲突的训练目标，以避免参数集过大。将这一扩展应用于多目标优化方法，将是后续文章的研究内容。

参考文献

[1] M. Raissi，P.Perdikaris，G. Karniadakis，物理信息神经网络：一种用于求解非线性偏微分方程正问题和逆问题的深度学习框架，《计算物理学杂志》378（2019）686–707。doi:10.1016/j.jcp.2018.10.045。

[2] J. Blechschmidt，O. Ernst，利用神经网络求解偏微分方程的三种方法——综述，GAMM-Mitteilungen 44（2）（2021年）e202100006。

[3] J. Hoffer、A. Ofner、F. Rohrhofer 等：《理论启发的机器学习——迈向知识与数据的协同》，《焊接世界》（2022年）。
doi:10.1007/s40194-022-01270-z。

[4] 罗伯特·科赫研究所，最后访问日期：2022年5月6日（2022年）。 [链接]。网址：https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Virologische_Basisdaten.html;jsessionid=960297181B52EB351E833DD09CD96CAB.internet081?nn=13490888#doc14716546bodyText6

[5] Ourworldindata.de，最后访问时间：2022年5月5日（2022年）。[链接]。网址：https://ourworldindata.org/coronavirus

[6] J. Malinzi、S. Gwebu、S. Motsa，使用物理信息神经网络确定新冠肺炎动态，Axioms 11(3)（2022年）121。doi:10.3390/axioms11030121。

[7] E. Kharazmi、M. Cai、X. Zheng、G. Lin、G. Karniadakis，使用物理信息神经网络研究整数阶和分数阶流行病学模型的可辨识性与可预测性，Nature Comput. Sci. 1(11)（2021年）744–753。doi:10.1101/2021.04.05.21254919。

[8] G. Pang、L. Lu、G. E. Karniadakis，分数阶物理信息神经网络，SIA M J. Sci. Comput. 41(4)（2019年）A2603–A2626。doi:10.1137/18M1229845。

[9] M. Cai、G. E. Karniadakis、C. Li，奥密克戎变异株新冠肺炎动态的分数阶SEIR模型与数据驱动预测，《Chaos》32（7）（2022年）071101。doi:10.1063/5.0099450。

[10] F. Rohrhofer、S. Posch、B. Geiger，物理信息神经网络的帕累托前沿，arXiv 预印本 2105.00862（2021年）。

[11] X. Jin, S. Cai, H. Li, G. E. Karniadakis，NSFnets（纳维–斯托克斯流网络）：用于不可压缩纳维–斯托克斯方程的物理信息神经网络，《计算物理学杂志》426 (2021年) 109951。doi:10.1016/j.jcp.2020.109951。

[12] S. Wang, Y. Teng, P. Perdikaris，理解并缓解物理信息神经网络中的梯度流病态，SIAM J. Sci. Comput. 43 (5) (2021年) A3055–A3081。doi:10.1137/20M1318043。[13] S. Maddu, D. Sturm, C. M¨uller, I. Sbalzarini，逆狄利克雷加权使物理信息神经网络的训练更加可靠，Machine Learning: Sci. Techn. 3 (2021年) 015026。doi:10.1088/2632-2153/ac3712。[14]J.-A. D´esid´eri，多目标优化的多梯度下降算法（MGDA），Compt. Rend. Math. 350 (2012) 313–318。doi:10.1016/j.crma.2012.03.014。[15] J. Fliege, B. Svaiter，多准则优化的最速下降法，Math. Meth. Oper. Res. 51 (3) (2000) 479–494。doi:10.1007/s001860000043。[16] O. Sener, V. Koltun，将多任务学习视为多目标优化，CoRR (2018)。arXiv:1810.04650v2。[17]S. Liu, L. Vicente，用于多目标优化的随机多梯度算法及其在监督式机器学习中的应用，Annals Oper. Res. (2021年)。doi:10.1007/s10479-021-04033-z。[18] L. McClenny, U. Braga-Neto，采用软注意力机制的自适应物理信息神经网络，arXiv预印本 2009.04544 (2020)。[19] A. F. Psaros, K. Kawaguchi, G. E. Karniadakis，元学习PINN损失函数，《计算物理学杂志》458 (2022年) 111121。doi:10.1016/j.jcp. 2022.111121。[20]M. Reiners, K. Klamroth, F. Heldmann, M. Stiglmayr，通过多目标学习方法进行权重剪枝，构建有效且稀疏的神经网络，Comput. Oper. Res. 141 (2022年) 105676。doi:10.1016/j.cor. 2021.105676。

[21] S. Berkhahn、M. Ehrhardt，物理信息神经网络：模拟新冠肺炎感染和住院情景，《Adv. Cont. Discr. Mod. Theo. Appl.》2022 (61) (2022)。doi:10.1186/s13662-022-03733-5。[22]W. Kermack、A. McKendrick，对流行病数理理论的贡献，《Bull. Math. Bio.》53 (1) (1991) 700–721。doi:10.1016/ S0092-8240(05)80040-0。[23]S. Treibert、H. Brunner、M. Ehrhardt，一种用于预测新冠肺炎动态的 SVICDR 模型非标准有限差分格式，《Math. Biosci. Engrg.》19 (2022) 1213–1238。doi:10.3934/mbe.2022056。[24] 罗伯特·科赫研究所，最后访问日期：2022年3月29日 (2022)。 [链接]。URL https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Fallzahlen_Kum_Tab.html[25] 罗伯特·科赫研究所，最后访问日期：2022年3月29日 (2022)。 [链接]。URL https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Klinische_Aspekte.html[26] 新冠肺炎疫苗接种仪表板，最后访问日期：2022年5月6日 (2022)。 [链接]。URL https://impfdashboard.de/daten[27] M. Martcheva，《数理流行病学导论》，第1版，Springer，2015年。doi:10.1007/978-1-4899-7612-3。[28] 罗伯特·科赫研究所，最后访问日期：2022年3月29日 (2022)。 [链接]。URL https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html[29] 德国联邦统计局，最后访问日期：2022年1月10日 (2022)。[链接]。URL https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/_inhalt.html

[30] R. E. Mickens，非线性反应-对流方程有限差分模型的精确解：对数值分析的启示，《J. Diff. Eqs. Appl.》9 (11) (2003) 313–325。doi:10.1080/1023619031000146959。[31] R. E. Mickens，《非标准有限差分格式的应用》，World Scientific (2000)。[32] M. Maamar、M. Ehrhardt、L. Tabharit，一种用于寨卡病毒传播时间分数阶模型的非标准有限差分格式，IMACM 预印本 22/21 (2022)。[33] G. E. Karniadakis、Y. Kevrekidis、L. Lu、P. Perdikaris、S. Wang、L. Yang，物理信息机器学习，《Nat. Rev. Phys.》3 (2021) 422–440。doi: 10.1038/s42254-021-00314-5。[34] S. Cuomo、V. S. di Cola、F. Giampaolo、G. Rozza、M. Raissi、F. Piccialli，通过物理信息神经网络实现科学机器学习：现状与未来方向 (2022)。doi:10.48550/ARXIV.2201.05624。[35]K. D. Olumoyin、A. Q. M. Khaliq、K. M. Furati，针对具有变化的缓解措施和传播率的无症状 COVID-19 模型的数据驱动深度学习算法，《Epidemiologia》2 (4) (2021)。doi:10.3390/epidemiologia2040033。[36] S. Shaier、M. Raissi、P. Seshaiyer，利用 DINNs（疾病信息神经网络）预测传染病传播的数据驱动方法 (2021)。doi:10.48550/ARXIV.2110.05445。[37] V. Grimm、A. Heinlein、A. Klawonn、M. Lanser、J. Weber，利用物理信息神经网络估计数理流行病学中 SIR 和 SEIR 模型的时变接触率，《Electr. Trans. Numer. Anal.》56 (2022) 1–27。doi:10.1553/etna_vol56s1。[38] A. G. Baydin、B. A. Pearlmutter、A. A. Radul、J. M. Siskind，机器学习中的自动微分：综述，《J. Mach. Learn. Res.》18 (2017) 5595–5637。

[39] Ben Moseley，最后访问时间：2022年12月12日（2021年）。 [链接]。网址：https://github.com/benmoseley/harmonic-oscillator-pinn[40] S. Ruder，《梯度下降优化算法概述》（2016）。 doi:10.48550/ARXIV.1609.04747。[41] M. Ehrgott，《多准则优化》，第2版，Springer，2005年。[42] A. Przybylski、K. Klamroth、R. Lacour，《一种用于多目标混合整数线性规划的简单有效的二分搜索算法》，arXiv预印本 1911.08937（2019年）。[43] I. Das、J. Dennis，《深入探讨在多准则优化问题中通过最小化目标函数加权和生成帕累托集的缺点》，Struct. Optim. 14（1997年）63–69。doi:10.1007/BF01197559。[44] K. Klamroth、K. Miettinen，《在多准则优化中整合近似方法与交互式决策》，Operations Research 56（2008年）222–234。doi:10.1287/opre.1070.0425。[45]M. Jagan、M. deJonge、O. Krylova、D. Earn，《快速估算时变传染病传播率》，PLoS Comput. Biol. 16（9）（2020年）e1008124。 doi:10.1371/journal.pcbi.1008124。[46]J. Yu、L. Lu、X. Meng、G. Karniadakis，《用于正向与逆偏微分方程问题的梯度增强物理信息神经网络》，Comput. Meth. Appl. Mech. Engrg. 393（2022年）114823。 doi:10.1016/j.cma.2022.114823。[47] T. Nagel、M. F. Huber，《用于系统辨识的 Kalman-Bucy信息神经网络》，arXiv预印本 arXiv:2210.03424（2022年）。[48] L. Yang、X. Meng、G. E. Karniadakis，《B-PINN：针对含噪声数据的正向与逆偏微分方程问题的贝叶斯物理信息神经网络》，《计算物理学杂志》425（2021年）109913。doi:10.1016/j.jcp.2020.109913。