面向过程优化的物理信息分段线性神经网络

Ece Serenat Koksal[a], Erdal Aydin[a,b,*]

[a]化学与生物工程系，科奇大学，土耳其伊斯坦布尔 *34450*[b]科奇大学 *TUPRAS* 能源中心（*KUTEM*），科奇大学，土耳其伊斯坦布尔，*34450*

摘要

由于现实过程十分复杂，构建第一性原理模型通常是一项困难且耗时的任务。另一方面，数据驱动建模，尤其是神经网络模型，往往面临过拟合以及缺乏有用且高质量数据等问题。与此同时，将训练好的机器学习模型直接嵌入优化问题，已成为一种有效且先进的代理优化方法；通过物理信息训练，还能进一步提升其性能。本研究提出在分段线性神经网络模型中融入物理信息，以改进嵌入神经网络模型的优化问题。除采用广受认可且天然具有分段线性特性的整流线性单元（ReLU）激活函数外，本研究还提出双曲正切激活函数的分段线性近似，以拓宽适用范围。研究考察了三个案例的优化问题：混合过程、工业蒸馏塔和原油塔。对于所有案例，经过物理信息训练的神经网络所得到的最优结果都更接近全局最优性。最后，这些优化问题所需的 CPU 时间远短于标准优化所需的时间。

关键词：机器学习、人工神经网络、物理信息神经网络、分段线性近似、混合整数线性规划。

* 通讯作者。电子邮箱：eaydin@ku.edu.tr（E. Aydin）。

1. 引言

人工神经网络是一种数学表达式，能够高效地概括过程的输入—输出关系，其灵感来源于大脑神经元连接。经典神经网络通常是完全由数据驱动的黑箱模型，其中不包含过程的第一性原理模型。尽管具有黑箱特性，近年来人工神经网络仍被广泛应用于文献研究中，因为在有数据可用时，它们尤其适用于复杂系统的建模。

化学工程系统的复杂性催生了对数据驱动模型的需求，尤其是人工神经网络，因为它们能够逼近非线性特性。然而，机器学习应用的黑箱特性、数据缺失或质量低下，以及过拟合等问题，都使得人们需要采用混合模型。例如，在间歇生产和半间歇生产等化学工程应用中，由于过程具有动态特性，数据有限且带有噪声（Thebelt 等，2022）。即使在连续过程中，许多与质量相关的中间产品规格也无法在线测量，从而引发了许多与在线优化和控制相关的问题。

近年来，物理信息混合模型已成为应对上述问题的一种解决方案。Stiasny 等人（2021年）将隐式龙格－库塔积分方案直接融入动态电力系统的物理信息人工神经网络训练模型。Nellikkath 和 Chatzivasileiadis（2021a） 将物理信息约束引入直流最优潮流模型，以保障安全并最大限度减少最坏情况违约。Nellikkath 和 Chatzivasileiadis（2021b）的另一项研究 将交流潮流方程引入神经网络，并减少最坏情况违约。Masi 等人（2021年）构建了一种基于热力学的人工神经网络模型，通过在网络中编码两条热力学定律来取代本构定律，从而预测材料响应。Pun 等人（2019年）将原子间键合引入

机器学习势，以提升其可迁移性并实现更好的原子级模拟。Haghighat & Juanes（2021年）使用 Python 开发了一个名为 SciANN 的工具箱，可将科学计算嵌入人工神经网络。HybridML 是另一个开源工具箱，可用于构建包含微分方程的混合模型。该工具箱介绍了一个案例研究，其中对 COVID-19 在德国联邦州的传播进行了建模（Merkelbach 等，2022）。

对于人工神经网络，双曲正切（tanh）、线性整流单元（ReLU）和 S形函数常用作隐藏层中的激活函数。线性整流单元由两个线性单元组成，在原点处存在不连续性，因此天然具有分段线性特性。ReLU 的分段线性特性使其适用于混合整数线性规划（混合整数线性规划）优化。Grimstad & Andersson（2019）展示了采用 ReLU 激活函数的神经网络的混合整数线性规划公式。Tsay 等（2021）提出了用于基于 ReLU 的神经网络的混合整数线性规划公式，在模型紧致性和规模之间取得平衡。Katz 等（2020a）将深度学习模型与多参数规划结合起来，并使用 ReLU 作为激活函数。该公式作为混合整数线性规划求解，从而降低了计算负担。Di Martino 等（2022）将经过 ReLU 训练的神经网络重新表述为用于过程优化的混合整数线性规划模型，并用该模型推导两个不同目标之间的 Pareto 前沿。Katz 等（2020b）将基于 ReLU 的神经网络与显式模型预测控制相结合，以控制太阳能场。Yang & Bequette（2021）采用基于 ReLU 的输入凸神经网络，将基于优化的控制问题转化为凸问题。通过凸形式，他们提升了控制器的鲁棒性和性能。Ceccon 等（2022）开发了优化与机器学习工具包（OMLT），该工具包利用预训练机器学习模型构建优化公式。reluMIP 工具箱使用 Gurobi 求解基于 ReLU 训练的人工神经网络模型的混合整数线性规划优化问题（Lueg 等，2021）。

双曲正切函数（tanh）可能对高度非线性过程具有更好的近似能力，也更有可能带来更好的训练和测试性能。对于双曲正切激活函数，Schweidtmann & Mitsos（2019）提出了缩减空间公式和 tanh 的松弛形式；针对混合整数非线性全局优化，他们采用基于 McCormick 的算法降低维度，并使用自研求解器 MaiNGO。对于非线性函数的分段线性近似，D’Ambrosio 等人（2010）提出了三种不同的公式，即一维方法、三角形方法和矩形方法，并讨论了各方法在近似质量与计算开销之间的权衡。Vielma（2015）建议通过引入辅助变量并采用基于线性规划（LP）的技术来构造混合整数线性规划公式。Sridhar 等人（2013）采用标准混合整数规划公式的改进版本来表示分段线性函数，并启用一个指示变量。Sildir & Aydin（2022）建议采用 3、5 和 7 段的双曲正切线性形式进行凸训练，并展示了这类近似能够获得相当紧致的下界，同时在训练和测试中的人工神经网络性能相近。

本研究将物理信息引导训练的分段线性神经网络与过程优化相结合：在将优化问题表述为混合整数线性规划（MILP）之前，先对非线性激活函数的分段线性形式进行近似。本文的主要贡献如下：(*i*) 将基于神经网络的混合整数线性规划优化研究扩展到不局限于 ReLU 激活函数的情形；(*ii*) 比较标准神经网络与物理信息神经网络在过程优化中的性能；(*iii*) 应用物理信息引导训练的模型，使用分段线性近似进行优化。第 2 节介绍训练与优化流程背后的方法和数学原理。第 3 节包含三个案例情景，用于探究物理信息神经网络在优化中的优势。各小节还探讨了不同的训练算法和优化公式，包括针对双曲正切函数的分段公式。结论见第 4 节。

2. 方法

2.1. 物理信息人工神经网络：训练

前馈全连接人工神经网络可表示为：

$$\hat{y} = f_1(Af_2(Bu + C) + D) \tag{1}$$

其中，f_1 和f_2 分别为输出层和隐藏层的激活函数，A和B为权重矩阵，C和D为偏置向量，u和y分别为输入向量和预测输出向量。

权重和偏置通过非线性优化进行估计，这一过程称为训练，其目标函数是最小化均方误差（MSE），如下所示：

$$min_{A,B,C,D} \frac{1}{N}\sum_{i=1}^{N}(f_1(Af_2(Bu_i + C) + D) - y_i)^2$$

$$s.t. \tag{2}$$

$$LW \leq A, B, C, D \leq UW$$

$$LH \leq f_2(Bu + C) \leq UH$$

其中，y_i 为第 i^{th} 个数据点，N为训练中的样本数，LW和UW分别为权重和偏置项的下界和上界，LH和UH分别为隐藏层输出的下界和上界。

为在训练中引入物理信息项，可以将损失函数设为双目标，同时最小化误差和物理项；也可以加入带有适当边界的额外约束。训练的双目标损失函数可写为：

$$min_{A,B,C,D}\frac{1}{N}\sum_{i=1}^{N}(f_1(Af_2(Bu_i+C)+D)-y_i)^2+(\alpha p)$$

$$s.t. \tag{3}$$

$$LW \le A,B,C,D \le UW$$

$$LH \le f_2(Bu+C) \le UH$$

其中，a 和 p 分别为标量权重值和物理函数项。当 p 作为约束引入时，训练问题（有时称为物理约束训练）可表示如下：

$$min_{A,B,C,D}\frac{1}{N}\sum_{i=1}^{N}(f_1(Af_2(Bu_i+C)+D)-y_i)^2$$

$$s.t.$$

$$\alpha p \le U \tag{4}$$

$$LW \le A,B,C,D \le UW$$

$$LH \le f_2(Bu+C) \le UH$$

其中，U表示与过程物理特性相关的某个上界。

由于额外的约束可能会增加非线性优化的复杂度，因此可以采用热启动将模型分为两个部分进行训练。首先，在不添加约束的情况下获得权重和偏置参数，以确定初始猜测值，这就是标准训练。然后，在第二部分中引入约束，并找到最优的权重和偏置参数，如下所示：

$$min_{A,B,C,D}z=\frac{1}{N}\sum_{i=1}^{N}(f_1(Af_2(Bu_i+C)+D)-y_i)^2 \tag{5a}$$

$$s.\ t.$$

$$LW \leq A, B, C, D \leq UW$$

$$LH \leq f_2(Bu + C) \leq UH$$

$$min_{A,B,C,D} z = \frac{1}{N}\sum_{i=1}^{N}(f_1(Af_2(Bu_i + C) + D) - y_i)^2$$

$$s.t. \tag{5b}$$

$$z \leq Z$$

$$\alpha p \leq U$$

其中，Z 是 MSE z 的上界。

2.2. 利用物理信息训练神经网络进行优化

标准过程优化的目标是在第一性原理过程模型和物理约束的条件下，寻找能够使选定输出最小化的过程最优输入。由于详细且基于物理的过程模型构建成本高昂，并且需要深入了解过程，因此在有过程数据可用时，代理优化是一种可行的替代方案。相应地，可以将物理信息神经网络模型作为等式约束嵌入优化问题中，其中包括训练后的神经网络模型的权重和偏置参数的最优值。这样，相比标准机器学习模型，物理信息机器学习模型，尤其是物理信息神经网络模型，或许能够提升代理优化的性能。

对于神经网络模型，使用 ReLU 激活函数并将训练后的神经网络模型直接嵌入优化问题中，利用 Big-M 公式化后，得益于 ReLU 的特性，可形成凸优化问题。在这种情况下，物理信息

训练和优化步骤可使用标准求解器和工具箱轻松完成（Abadi 等，2016；Ceccon 等，2022；Lueg 等，2021）。然而，当模型使用非线性的双曲正切作为激活函数进行训练时，利用训练后的神经网络模型寻找最优解颇具挑战，因为由此得到的公式是一个非凸非线性规划问题。Karuppiah 与 Grossmann（2006）提出了一种用于求解非凸非线性规划公式的空间分支定界算法，其中使用分段线性估计器近似非凸项，以获得全局最优解的下限。Schweidtmann 与 Mitsos（2019）提出了采用凸包络与凹包络的缩减空间公式，以严谨的方式通过确定性全局优化解决这一问题。

在本研究中，为确保代理优化问题具有凸性，采用线性分段近似双曲正切函数。因此，可以使用标准工具箱和原始双曲正切激活函数重新训练物理信息神经网络。随后，使用双曲正切函数的分段近似进行优化。事实上，在相同的设置下，也可以使用 S形函数和其他类型的激活函数。值得庆幸的是，作者已经证明，此类近似通常能得到紧致的松弛和准确的表示（Sildir & Aydin，2022）。给定 -4、-1、 +1 和 +4 为双曲正切函数（tanh）的断点，tanh 的三段和五段线性近似如下所示（D'Ambrosio 等，2010）：

$$\tanh(x)_{3\,pcs} \approx \begin{cases} 0.08x - 0.68 & if\ -4 \le x \le -1 \\ 0.76x & if\ -1 \le x \le +1 \\ 0.08x + 0.68 & if\ -1 \le x \le +4 \end{cases} \tag{6}$$

$$\tanh\,(x)_{5\,pcs} \approx \begin{cases} 0.018x - 0.93 & if\ -4 \le x \le -2 \\ 0.20x - 0.56 & if\ -2 \le x \le -1 \\ 0.76x & if\ -1 \le x \le +1 \\ 0.20x + 0.56 & if\ +1 \le x \le +2 \\ 0.018x + 0.93 & if\ +2 \le x \le +4 \end{cases} \tag{7}$$

其中 $\tanh(x)_{3\ pcs}$ 和 $\tanh(x)_{5\ pcs}$ 分别表示 tanh 的三段近似和五段近似。 图 1。展示了 tanh 的三段和五段近似。五段形式的近似效果更好，因为它具有更多断点和线性分段。然而，当将问题建模为混合整数线性规划（MILP）时，变量数量会随线性分段数量的增加而增加。

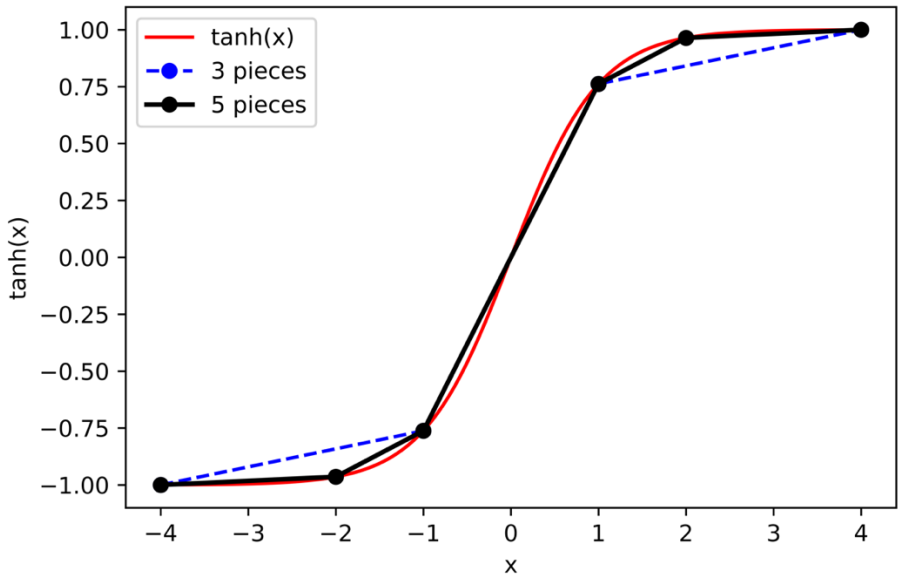

图 1。 双曲正切的近似。

最后，基于双曲正切的近似物理信息神经网络模型可以利用（Sildir & Aydin, 2022; Vielma, 2015）中的概念，构建为一个混合整数线性优化问题，以寻找使训练好的模型的目标输出最小化的最优输入，如下所示：

$$\min_{u} y = f_1(A^* f_2(B^* u + C^*) + D^*)$$

$$s.t.$$

$$(-4\lambda_1) + (-\lambda_2) + (\lambda_3) + (4\lambda_4) = z_2 \tag{8}$$

$$(-\lambda_1) + (-0.76\lambda_2) + (0.76\lambda_3) + (\lambda_4) = f_2'(z_2)$$

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$$

$$\gamma_1 + \gamma_2 + \gamma_3 = 1$$

$$d_1 + d_2 + d_3 + d_4 = 1$$

$$(-4d_1) + (-d_2) + (d_3) + (4d_4) = z_1$$

$$(-d_1) + (-0.76d_2) + (0.76d_3) + (d_4) = f_1'(z_1)$$

$$-1 \leq f_1'(z_1) \leq 1$$

$$-1 \leq f_2'(z_2) \leq 1$$

$$-4 \leq z_1 \leq 4$$

$$-4 \leq z_2 \leq 4$$

$$\lambda_1 \leq \gamma_1$$

$$\lambda_2 \leq \gamma_1 + \gamma_2$$

$$\lambda_3 \leq \gamma_2 + \gamma_3$$

$$\lambda_4 \leq \gamma_3$$

$$d_1 \leq \beta_1$$

$$d_2 \leq \beta_1 + \beta_2$$

$$d_3 \leq \beta_2 + \beta_3$$

$$d_4 \leq \beta_3$$

$$\lambda_k \geq 0 \quad \forall k \in \{1, \ldots 4\}$$

$$\delta_k \geq 0 \quad \forall k \in \{1, \ldots 4\}$$

$$0 \leq \gamma_j \leq 1 \quad \forall j \in \{1,2,3\}$$

$$0 \leq \beta_j \leq 1 \quad \forall j \in \{1,2,3\}$$

$$\gamma_j \in Z \; \forall j \in \{1,2,3\}$$

$$\beta_j \in Z \; \forall j \in \{1,2,3\}$$

其中，l_j 和d_j 是非负辅助变量，k是单层中非负辅助变量的数量，g_j 和b_j 是与断点相关的二元变量，j是

近似的线性分段数量，z_1 和z_2 分别是输出层和隐藏层的输入，其定义域为 -4 至 +4；f_1' 和 f_2' 分别是输出层和隐藏层的双曲正切近似函数，其定义域为 -1 至 +1；A^*、B^*、C^* 和D^* 分别是训练好的模型的权重和偏置参数。

最后，式 8 所示的问题也可以通过特殊有序集 II 型（SOS2）变量表述如下（Sridhar 等人，2013）：

$$
\begin{aligned}
&\min_u y = f_1(A^* f_2(B^* u + C^*) + D^*) \\
&s.t. \\
&\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1 \\
&(-4\lambda_1) + (-\lambda_2) + (\lambda_3) + (4\lambda_4) = z_1 \\
&(-\lambda_1) + (-0.76\lambda_2) + (0.76\lambda_3) + (\lambda_4) = f_1'(z_1) \\
&\delta_1 + \delta_2 + \delta_3 + \delta_4 = 1 \\
&(-4\delta_1) + (-1\delta_2) + (1\delta_3) + (4\delta_4) = z_2 \\
&(-1\delta_1) + (-0.76\delta_2) + (0.76\delta_3) + (1\delta_4) = f_2'(z_2) \\
&0 \le \lambda_k \le 1, k = 1,2,3,4 \\
&0 \le \delta_k \le 1, k = 1,2,3,4 \\
&-1 \le f_1'(z_1) \le 1 \\
&-1 \le f_2'(z_2) \le 1 \\
&-4 \le z_1 \le 4 \\
&-4 \le z_2 \le 4 \\
&\lambda \; and \; \delta \; are \; SOS2 \; variables.
\end{aligned}
\tag{9}
$$

该方法的示意图见图2，其中u^* 表示最优输入集。

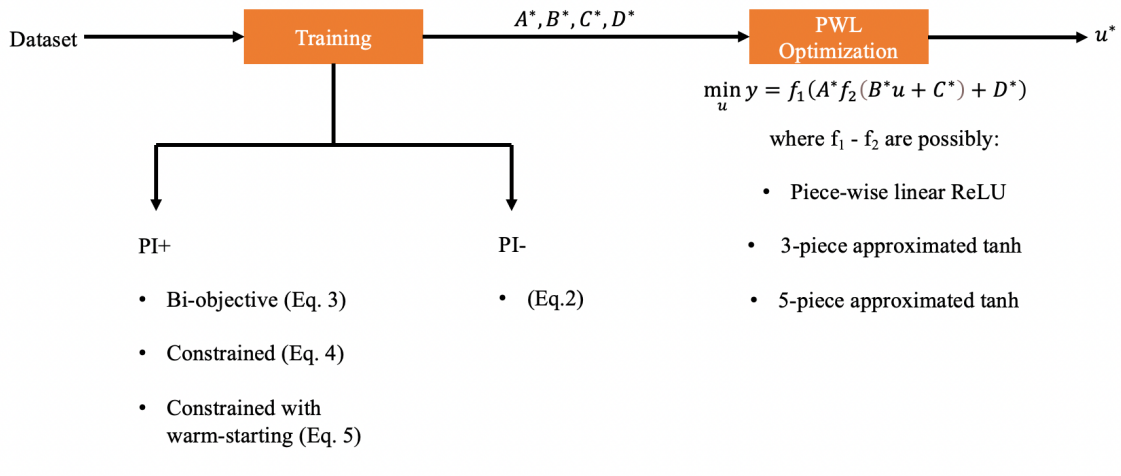
Dataset
Training
A^*, B^*, C^*, D^*
PWL Optimization
u^*
$\min_u y = f_1(A^* f_2(B^* u + C^*) + D^*)$
where f_1 - f_2 are possibly:
• Piece-wise linear ReLU
• 3-piece approximated tanh
• 5-piece approximated tanh
PI+
• Bi-objective (Eq. 3)
• Constrained (Eq. 4)
• Constrained with warm-starting (Eq. 5)
PI-
• (Eq.2)

图2。 方法流程图。

3. 结果与讨论

3.1. 案例研究 1：一个简单的混合过程

两个流股中两种互溶液体进行混合的混合过程如图 3所示。在此，流股由组分 A 和 B 组成的混合物构成，且不发生反应。输入x_1, x_2, w_1, 和w_2 分别表示流股 1 中组分 A 的摩尔分数、流股 2 中组分 A 的摩尔分数、流股 1 的摩尔流率和流股 2 的摩尔流率。输出由 x和 w表示，分别为出口流股中组分 A 的摩尔分数和总流率。人工神经网络的输入和输出向量如下所示：

$$u = \begin{bmatrix} x_1 \\ x_2 \\ w_1 \\ w_2 \end{bmatrix}$$

$$y = \begin{bmatrix} x \\ w \end{bmatrix}$$

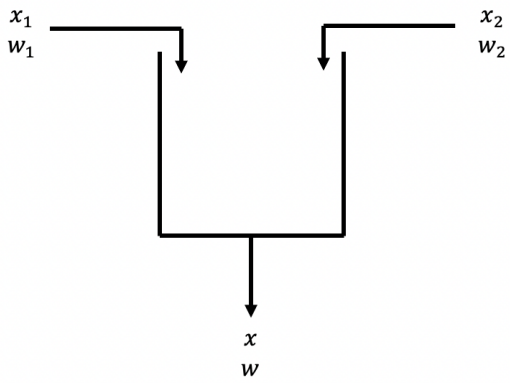

图 3。 混合过程。

该过程的第一性原理模型方程如下：

$$\begin{gathered} x_1 w_1 + x_2 w_2 = xw \\ w_1 + w_2 = w \\ 0.1 \le x_1 \le 0.5 \\ 0.5 \le x_2 \le 1 \\ 0 \le w_1, w_2 \le 50 \end{gathered} \tag{10}$$

基于式（10）中的模型，使用 Matlab 中的“rand”命令对输入数据进行采样，生成大小为 100 的合成数据集。数据归一化到 -1 至 1 之间，使用“awgn”函数添加白噪声，并通过“randperm”打乱数据。80% 的数据用于训练。除输出层外，模型还使用一个包含 5 个神经元的隐藏层，并选择双曲正切作为非线性激活函数。训练流程如下：

(i) 无物理信息（PI-）情形，其中目标函数是使均方误差最小化；(ii) 物理信息驱动（PI+）情形，通过使用式 10 中的第一项（即 A 的组分质量衡算），将物理项嵌入 PI+ 人工神经网络中。该项的计算见式 11：

$$p = \sum_{i=1}^{N} \left(\hat{y}_{1,i} \hat{y}_{2,i} - u_{i,1} u_{i,3} - u_{i,2} u_{i,4} \right)^2 \tag{11}$$

对于无物理信息（PI-）和物理信息驱动（PI+）两种情况，分别选择公式2和公式3作为目标函数。在优化部分，目标是最小化出口流股的摩尔分数，同时要求流股1和2的最小流量为25个单位。考虑到当 x 的定义域为 -4 至 +4 时，双曲正切(x) 的取值介于 -1 和 +1 之间，采用精确形式的双曲正切(x) 的代理优化问题可表示为方程（12）：

$$\begin{gathered} \min_{u} y_1 = f_1(A^* f_2(B^* u + C^*) + D^*) \\ s.t. \\ -1 \le f_1'(z_1) \le 1 \\ -1 \le f_2'(z_2) \le 1 \\ 25 \le u_3 \end{gathered} \tag{12}$$

$$25 \leq u_4$$

$$-4 \leq z_1 \leq 4$$

$$-4 \leq z_2 \leq 4$$

针对两种不同的情况进行优化：

(i) 如方程（12）所示，优化问题被表述为非线性规划，其中激活函数使用精确的双曲正切，并采用 BARON 作为非凸全局求解器。这是采用精确双曲正切形式的代理优化，表示为 非线性规划$_{tanh}$。(ii) 通过将双曲正切的三段和五段近似用作训练好的神经网络的激活函数，将方程（12）中的问题转化为混合整数线性规划，并使用 CPLEX 求解。相应的公式分别表示为 混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$ 和 混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$。此案例研究中双曲正切三段近似的混合整数线性规划公式见方程（13）：

$$\begin{aligned}
&\min_u y_1 = f_1(A^* f_2(B^* u + C^*) + D^*) \\
&s.t. \\
&\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1 \\
&(-4\lambda_1) + (-\lambda_2) + (\lambda_3) + (4\lambda_4) = z_1 \\
&(-\lambda_1) + (-0.76\lambda_2) + (0.76\lambda_3) + (\lambda_4) = f_1'(z_1) \\
&\delta_1 + \delta_2 + \delta_3 + \delta_4 = 1 \\
&(-4\delta_1) + (-\delta_2) + (\delta_3) + (4\delta_4) = z_2 \\
&(-\delta_1) + (-0.76\delta_2) + (0.76\delta_3) + (\delta_4) = f_2'(z_2) \\
&0 \leq \lambda_k \leq 1, k = 1,2,3,4 \\
&0 \leq \delta_k \leq 1, k = 1,2,3,4 \\
&-1 \leq f_1'(z_1) \leq 1
\end{aligned} \tag{13}$$

$$-1 \le f_2'(z_2) \le 1$$

$$0 \le u_3 \le 1$$

$$0 \le u_4 \le 1$$

$$-4 \le z_1 \le 4$$

$$-4 \le z_2 \le 4$$

$$\lambda \; and \; \delta \; are \; SOS2 \; variables。$$

对于 PI- 和 PI+，每种情况均报告了 5 次运行的训练和测试均方误差、输入集、目标函数以及求解器的 CPU 时间，见表1。进行五次试验的原因是，人工神经网络模型由局部求解器训练。由于训练问题具有非凸性，训练问题可能收敛到不同的局部最优解，因此训练后的神经网络可能彼此不同。不过，最优值之间的差异，尤其是在物理信息案例中，应当较小，因为其可靠性有所提高。作者还提出了一种用于神经网络训练和特征选择的凸训练方法（Sildir & Aydin，2022）。

表1。

案例研究 1 的训练、测试和优化性能。

训练条件	*Run #*	训练 *MSE*	*Test MSE*	公式	x_1	x_2	w_1	w_2	x	w	*CPU* 时间（秒）
PI-	1	0.0062	0.0100	非线性规划$_{tanh}$	0.12	0.5	35.9	25	0.28	70.0	10.2
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.29	0.5	25	25	0.48	44.2	0.52
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.26	0.5	47.9	25	0.30	62.9	0.88
	2	0.0022	0.0044	非线性规划$_{tanh}$	0.1	0.5	49.6	25	0.24	78.9	1.50
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.1	0.5	49.6	25	0.20	65.3	0.01
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.1	0.5	49.6	25	0.18	65.3	0.03
	3	0.0032	0.0072	非线性规划$_{tanh}$	0.1	0.5	38.5	25	0.29	68.5	7.15
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.1	0.5	50.0	50	0.35	91.5	0.26
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.1	0.5	42.4	25	0.27	79.3	0.53
	4	0.0033	0.0067	非线性规划$_{tanh}$	0.1	0.5	50	25	0.22	82.7	2.53
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.1	0.5	50	25	0.19	82.7	0.56
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.1	0.5	50	25	0.16	82.9	0.36
	5	0.0036	0.0068	非线性规划$_{tanh}$	0.1	0.5	50	25	0.29	81.0	6.71
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.1	1	41.9	25	0.30	61.7	0.69
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.1	0.57	25	25	0.28	52.2	0.49
PI+	1	0.0034	0.0048	非线性规划$_{tanh}$	0.1	0.5	50	25	0.24	79.9	1.07
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.1	0.5	50	25	0.21	64.9	0.34
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.1	0.5	50	25	0.21	65.4	0.54
	2	0.0060	0.0098	非线性规划$_{tanh}$	0.1	0.5	50	25	0.21	78.9	1.61
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.1	0.5	49.7	25	0.21	71.3	0.42
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.1	0.5	49.7	25	0.21	72.8	0.35
	3	0.0030	0.0043	非线性规划$_{tanh}$	0.1	0.5	50	25	0.23	79.9	1.72
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.1	0.5	50	25	0.20	63.7	0.01
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.1	0.5	50	25	0.21	78.0	0.50
	4	0.0064	0.0093	非线性规划$_{tanh}$	0.1	0.5	50	25	0.26	77.8	0.75
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.1	0.5	50	38.5	0.34	82.6	0.63
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.1	0.5	50	38.5	0.34	82.6	0.26
	5	0.0030	0.0043	非线性规划$_{tanh}$	0.1	0.5	50	25	0.23	79.9	1.61
				混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	0.1	0.5	50	25	0.20	63.7	0.03
				混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	0.1	0.5	50	25	0.21	78.0	0.89
全局				*NLP*	0.1	0.5	50	25	0.23	75	0.01

PI- 的平均训练均方误差和测试均方误差分别为 0.0037 和 0.0070，PI+ 的相应数值分别为 0.0043 和 0.0065。在 PI+ 情况下，由于数据中的噪声导致偏离质量守恒，训练性能有所下降。因此，将物理项引入

目标函数会使训练性能略有下降。另一方面，由于训练中的物理项能够提供更好的预测，测试性能得以提升。

对于优化部分，除了代理优化运行之外，还使用第一性原理模型将模型表述为非线性规划，并使用非凸全局求解器 BARON 求解。全局最优输入和输出列于表1的最后一行。全局解表明，要使出口流股的摩尔分数达到最小，浓缩物流的摩尔分数和流量x_1, x_2, 以及w_2, 应取最小值，而稀释物流的流量w_1, 应取最大值，这也可以用工程直觉来解释。

当 PI+ 情形被表述为非线性规划$_{tanh}$，并在代理优化中采用双曲正切的精确形式，且由非凸全局求解器 BARON 求解时，决策变量（输入）的取值总是收敛到全局最优解。另一方面，采用精确双曲正切激活函数的 PI- 训练好的神经网络的最优输入非线性规划$_{tanh}$可能会偏离全局最优解，例如表1中 PI- 情形的运行1和3表1所示。这表明，在嵌入神经网络模型的代理优化问题中，物理信息训练本身具有优势。

由于物理信息项基于组分质量衡算，而非总质量守恒，且人工神经网络使用噪声数据进行训练，因此输出物流的流量w偏差要显著得多，尤其是 PI- 情形的最后 3 次运行。考虑到数据的质量和数量有限，这一结果很有前景。事实上，当问题规模较大或数据质量较低时，与其增加神经元数量，使问题变得更加复杂并可能导致过拟合，不如使用物理信息神经网络来实现更可靠的代理优化。

另一个观察结果是，采用非线性规划$_{tanh}$公式化优化 PI+ 训练好的神经网络时，CPU 时间较短。不过请注意，对于更大的神经网络和深度学习，使用这类高度非线性激活函数的精确形式可能并不实用（Schweidtmann & Mitsos，2019）。因此，我们考察了采用 PI+ 分段线性神经网络的代理优化结果；这类网络可构成混合整数线性规划问题，也是本研究提出的另一项贡献。

采用双曲正切近似的物理信息驱动混合整数线性规划建模优化，其结果与采用精确双曲正切的物理信息驱动优化相比，偏离了全局最优解（非线性规划$_{tanh}$）。另一方面，正如预期，将优化建模为混合整数线性规划后，CPU 时间显著缩短。即使神经元数量较少，这些差异也可能很明显。采用精确双曲正切的公式具有非凸性，尤其在使用全局非凸求解器时，会增加 CPU 时间。另一方面，对双曲正切进行分段近似可将问题转化为混合整数线性规划。由于其凸性，CPU 时间大幅缩短。

该案例研究的问题规模较小，神经元数量也较少，因此全局解与混合整数线性规划解的比较较为直接。对于隐藏层数量较多的情况（深度学习）以及神经元数量较多的情况，求解非线性规划问题预计会更具挑战性，甚至可能无法求解；而对于采用物理信息训练神经网络的近似混合整数线性规划优化问题，情况则并非如此（Schweidtmann & Mitsos，2019）。因此，对于所研究的案例研究，采用近似双曲正切激活函数的混合整数线性规划公式，虽然会带来与全局解的微小偏差，但能显著加快求解速度。

除了混合整数线性规划公式在计算上的优势外，对近似的双曲正切激活函数进行物理信息训练，也有助于缩小代理优化与精确全局结果之间的差距。如表1所示，混合整数线性规划公式的偏差

相较于 NLP_{tanh} 和全局NLP，在无物理信息的 PI−案例中要显著得多。同时，对于 PI+，混合整数线性规划$_{5\text{-pcs-tanh}}$ 解比NLP_{tanh} 更接近混合整数线性规划$_{3\text{-pcs-tanh}}$，因为五段近似具有更多线性分段和断点，从而具备更强的近似能力。因此，使用更多分段来近似精确的双曲正切，即使只付出少量 CPU 时间增加的代价，也可能使与精确全局最优解之间的偏差小到可以忽略。

3.2. 案例研究2：工业蒸馏塔

出于专有信息保护的原因，本案例研究使用了 Tüpraş Kırıkkale炼油厂原油分馏塔的归一化数据。数据包含原油（CR）、液化石油气（LPG）、轻直馏石脑油（LSRN）、重直馏石脑油（HSRN）、轻柴油（LD）、重柴油（HD）、煤油（kero）和常压渣油（RSD）的流量数据。图4展示了该塔的示意图。

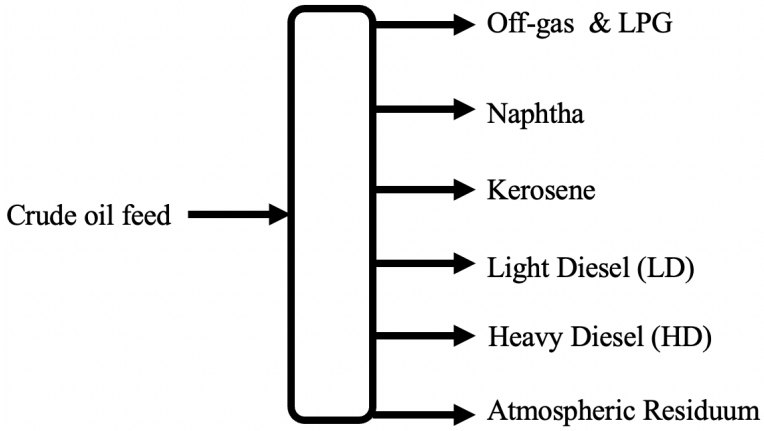

图4。原油塔示意图。

人工神经网络的输入选为原油流量和其他产品流量，残渣油除外；残渣油将作为人工神经网络的预测输出。人工神经网络的输入和输出向量如下：

$$u = \begin{bmatrix} F_{CR} \\ F_{LPG} \\ F_{LSRN} \\ F_{HSRN} \\ F_{kero} \\ F_{LD} \\ F_{HD} \end{bmatrix}$$

$$y = [F_{RSD}]$$

其中，F表示流量。

由于流量　范围不同，因此通过反归一化引入质量平衡方程　。物理项 p 定义如下：

$$p = \frac{\sum_{i=1}^{N}\left(u_{1,i} - \sum_{m=2}^{7} u_{m,i} - \hat{y}_i\right)}{N} \tag{14}$$

其中 y" 是残渣油流量的预测值。数据归一化至 -1 到 1 之间，神经网络通过两种不同方式进行训练：

（i）以均方误差作为目标函数进行最小化，如公式2所示，称为非物理信息驱动（PI-）情形；（ii）对均方误差和 p 项加权，并将损失函数构建为双目标形式，如公式3所示，称为物理信息驱动（PI+）情形。

该神经网络包含一个有五个神经元的隐藏层，隐藏层和输出层均采用双曲正切作为非线性激活函数。在 270 个数据点中，80% 用于训练，其余用于测试。假设塔入口处的原油流量达到最大值，并考虑到当决策变量区间为 [-4, +4] 时，tanh(x) 的取值介于 -1 和 +1 之间，通过以下代理优化问题求得使常压渣油流量最小的最优输入：

$$\min_{u} y = f_1(A^* f_2(B^* u + C^*) + D^*)$$

$$s.t.$$

$$-1 \le f_1'(z_1) \le 1$$

$$-1 \le f_2'(z_2) \le 1 \qquad (15)$$

$$u_1 = \eta_{CR,max}$$

$$-4 \le z_1 \le 4$$

$$-4 \le z_2 \le 4$$

其中，$h_{CR,\ max}$ 是最大归一化原油值，在本案例研究中等于 1。请注意，遗憾的是，本文未提供所研究蒸馏塔的详细第一性原理模型。不过，由于塔必须遵循质量守恒，在假设不存在泄漏的情况下，可基于质量平衡建立模型，以预测全局解，并将其表示为估计全局解（*EGS*）。基于质量平衡得到的全局解意味着，在原油量达到最大值时，其他产品也必须达到最大值，才能使常压渣油量最小。

方程 15 中的优化问题也通过对双曲正切使用 3 段和 5 段近似，构建为混合整数线性规划。混合整数线性规划公式 (混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$, 混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$) 以及方程 15 中采用精确双曲正切的非线性规划形式 (非线性规划$_{tanh}$)，分别由 CPLEX 和非凸全局求解器 BARON 求解。

由于训练中普遍存在非凸性问题，为避免次优训练的影响，该模型进行了五次训练，并分别针对这些训练样本进行优化。训练和测试性能见表2。

表2。案例研究2的训练和测试性能。

训练条件	*PI-*		*PI+*	
运行编号	训练均方误差	测试均方误差	训练均方误差	测试均方误差
1	2.12×10^{-5}	5.38×10^{-5}	7.56×10^{-6}	2.44×10^{-5}
2	3.83×10^{-5}	1.23×10^{-4}	1.50×10^{-5}	4.58×10^{-5}
3	2.04×10^{-5}	4.90×10^{-5}	1.13×10^{-5}	3.47×10^{-5}
4	3.54×10^{-4}	6.28×10^{-4}	1.48×10^{-5}	4.94×10^{-5}
5	2.16×10^{-5}	7.12×10^{-5}	1.31×10^{-5}	5.46×10^{-5}
平均值	9.11×10^{-5}	1.85×10^{-4}	1.23×10^{-5}	4.18×10^{-5}

在本案例研究的所有运行中，当在训练目标函数中加入加权物理项后，训练和测试性能均得到提升。对于每个训练好的模型，三种不同优化形式的结果见表3，其中η 表示变量的归一化值。

表3。案例研究2的最优输入。

物理条件	运行 公式		h_{LPG}	h_{LSRN}	h_{HSRN}	h_{kero}	η_{LD}	h_{HD}	h_{RSD}	*CPU time (s)*
PI-	1	非线性规划$_{tanh}$	1	1	1	1	1	1	0.21	0.76
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.23	0.51
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.12	0.53
	2	非线性规划$_{tanh}$	0.42	1	1	1	1	1	0.17	22.0
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.15	0.75
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.03	0.61
	3	非线性规划$_{tanh}$	-0.20	1	1	1	1	1	0.25	9.58
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	-0.20	1	1	1	1	1	-0.12	0.74
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	-0.20	1	1	1	1	1	0.10	0.56
	4	非线性规划$_{tanh}$	1	-0.10	1	1	-0.04	1	0.53	69.5
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	1	-0.41	1	1	0.16	1	0.67	0.68
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	1	-0.29	1	1	0.08	1	0.66	0.571
	5	非线性规划$_{tanh}$	1	0.65	1	1	1	1	0.25	1.71
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	1	0.65	1	1	1	1	0.56	0.42
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	1	0.65	1	1	1	1	0.22	0.56
PI+	1	非线性规划$_{tanh}$	1	1	1	1	1	1	0.19	1.53
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	1	-0.58	1	1	1	1	-0.17	0.69
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.20	0.52
	2	非线性规划$_{tanh}$	1	1	1	1	1	1	0.20	4.91
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	-1	1	1	0.57	1	1	0.17	0.51
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.17	0.44
	3	非线性规划$_{tanh}$	1	1	1	1	1	1	0.20	3.25
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	-0.34	1	1	1	1	1	0.26	0.40
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.18	0.51
	4	非线性规划$_{tanh}$	1	1	1	1	1	1	0.20	1.26
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	-1	1	1	0.94	1	1	0.15	0.58
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.17	0.61
	5	非线性规划$_{tanh}$	1	1	1	1	1	1	0.20	1.37
		混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.17	0.44
		混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$	1	1	1	1	1	1	0.22	0.48
EGS			1	1	1	1	1	1	0.20	0.00

当使用神经网络隐藏层中的精确双曲正切激活函数来构建优化问题时，如非线性规划$_{tanh}$，PI+ 训练好的模型始终收敛于估计的全局最优解

但 PI+ 的第1次运行中 h_{RSD} 存在轻微偏差，这与第一个案例研究类似。结果还表明，PI- 情形也有可能收敛到估计的全局解，如第1次运行所示。另一方面，PI- 的其他运行结果表明其缺乏一致性，而且 PI- 训练好的神经网络模型通常不太可能收敛到估计的全局解。因此可以认为，在物理信息无关代理优化情形下，训练对优化的影响要大得多，因为 PI+ 训练好的神经网络既能持续获得接近最优的结果，也表现出一致性。

除了采用 PI+ 后精度更高之外，尽管分段线性网络的近似会带来偏差，但在物理信息训练下，混合整数线性规划公式的结果仍接近非线性规划$_{tanh}$ 解。事实上，与用于代理优化的 PI- 分段线性神经网络相比，PI+ 分段线性网络的偏差甚至可以忽略不计。此外，CPU 时间表明，由于凸性，混合整数线性规划公式所需的计算时间远短于非线性规划$_{tanh}$。换言之，在 PI+ 训练下，近似能够产生紧致的松弛。此外，与具有更多线性分段和断点的混合整数线性规划$_{3\text{-}pcs\text{-}tanh}$ 相比，混合整数线性规划$_{5\text{-}pcs\text{-}tanh}$ 结果更接近全局最优解。因此，物理信息训练不仅能使优化结果更接近全局最优解，还能为规模更大的基于神经网络的混合整数线性规划优化带来计算优势；本案例研究也观察到了这一点。

3.3. 案例研究3：原油装置

原油装置（原油蒸馏装置）是炼油厂中的初始分离装置，用于将含有长碳链的原油转化为白色产品，主要包括液化石油气、石脑油、煤油、柴油和减压瓦斯油。从底部可获得常压渣油。该产品可用作沥青，或送往其他装置进行进一步裂化，例如减压蒸馏。在本案例研究中，采用 Alhajri等（2008）的模型，并进行了部分修改。模型中使用的切割点温度范围列于 表4。

表4。切割点温度范围。

产品	沸程 下限（F）	沸程 上界（F）
LPG	-48	-40
石脑油	230	380
煤油	330	520
柴油	420	630
VGO	620	1050

产品的切割点通过公式 16 求得:

$$Cut_s = \sum_{k=0}^{4} a_k (TE_{CDU,s})^k \tag{16}$$

其中，*s*表示原油蒸馏装置的产品，但不包括*RSD*（残渣油），$TE_{CDU,\ s}$ 是产品 *s*的切割点温度，而a_k 是表5中给出的常数表5（Alhajri 等人，2008年）。由于切割是累积的，*RSD*的切割值等于100。

表5。

切点方程的常数。

a_0	4.04
a_1	-0.047
a_2	3.25×10^{-4}
a_3	-2.84×10^{-7}
a_4	8.15×10^{-11}

每种产品的流量计算如下：

$$F_{CDU,s} = F_{CDU} \times \left(\frac{Cut_s - Cut_{s-1}}{100}\right) \tag{17}$$

在不同的原油流量和切割温度下，使用拉丁超立方抽样（Matlab 中的“lhsdesign”函数）创建了一个包含 1000 个样本的合成数据集，并使用模型方程获得这些输入对应的产品流量。人工神经网络和输出变量如下所示：

$$u = \begin{bmatrix} F_{CDU} \\ TE_{LPG} \\ TE_{Naphtha} \\ TE_{Kerosene} \\ TE_{Diesel} \\ TE_{VGO} \end{bmatrix} \qquad y = \begin{bmatrix} F_{LPG} \\ F_{Naphtha} \\ F_{Kerosene} \\ F_{Diesel} \\ F_{VGO} \\ F_{RSD} \end{bmatrix}$$

Diff采用不同的训练算法，并在表6中给出了它们的均方误差值。

表6。

案例研究 3 的训练和测试性能。

神经元数量位于隐藏层中	训练算法	隐藏层激活函数	训练条件	训练 *MSE*	*Test MSE*
10	Adam 优化器	tanh	PI-	0.000118	0.000125
			PI+	0.000128	0.000144
	Adam 优化器	ReLU	PI-	0.000519	0.000560
			PI+	0.000456	0.000496
	CONOPT	tanh	PI-	0.000020	0.000021
			PI+	0.000018	0.000018
15	Adam 优化器	tanh	PI-	0.000054	0.000061
			PI+	0.000063	0.000067
20	Adam 优化器	tanh	PI-	0.000047	0.000051
			PI+	0.000048	0.000051

表7总结了该案例研究的优化表现，其中采用了不同数量的神经元、训练算法和条件、激活函数及优化方法。

表7。

案例研究 3 的优化结果。

# of 隐藏层中的神经元层	训练算法 训练目标函数		隐藏层激活函数 *at* 训练	优化方法	*# of* 可行的运行次数 *of 5*	偏差比率 F_{RSD} 来自全局解	平均值 *SSE* $(x10^{-4})$
10	Adam 优化器	(公式2) (PI-)	tanh	3段双曲正切	4	0.08	5.3
		(公式3) (PI+)			4	0.06	4.8
		（式2）（PI-）		5段双曲正切	3	0.05	4.4
		（式3）（PI+）			4	0.04	4.0
		（式2）（PI-）	ReLU	relu MIP	5	0.00	9.3
		（式3）（PI+）			5	0.00	6.6
	CONOPT	（式2）（PI-）	tanh	3段双曲正切	3	0.13	38.2
		（式5）（PI+）			5	0.07	34.7
15	Adam 优化器	（式2）（PI-）	tanh	3段双曲正切	4	0.14	7.7
		（式3）（PI+）			4	0.08	3.2
		（式2）（PI-）		5段双曲正切	4	0.07	3.9
		（式3）（PI+）			4	0.04	2.2
20	Adam 优化器	（式2）（PI-）	tanh	3段双曲正切	4	0.16	7.9
		（式3）（PI+）			5	0.08	6.3
		（式2）（PI-）		5段双曲正切	4	0.10	3.9
		（式3）（PI+）			5	0.04	4.3

3.3.1. 双目标物理信息训练与分段线性优化 n

本案例研究使用 Keras 库内置的自定义损失函数来计算预测值与真实值（Chollet 等，2015）。物理项是蒸馏塔输入与输出流量之差的平方根，用于表示质量平衡的平方根：

$$p = \sqrt{\sum_{i=1}^{N} u_{1,i} - \sum_{j=1}^{6} \hat{y}_{i,j}} \tag{17}$$

其中，$u_{1,i}$ 是第 i^{th} 条数据的第一个人工神经网络输入，表示原油流量；而 $y''_{i,j}$ 是人工神经网络输出中第 i^{th} 条数据的第 j^{th} 列，表示蒸馏过程中获得的白色产品流量。

神经网络模型在 TensorFlow 中进行训练，采用单个隐藏层，分别使用 10、15 和 20 个神经元。采用 Adam 优化器作为训练算法（Kingma和Ba，2014年）。对于 10、15 和 20 个神经元，隐藏层的激活函数均采用双曲正切。每种情况均运行 5 次，并报告训练和测试的均方误差平均值，见 表6。

对于隐藏层中的每个神经元数量，采用物理信息训练时，训练的均方误差值可能会略有变化。训练阶段得到的权重和偏置参数将作为参数输入优化问题，其目标是最小化残渣油流量。通过这种方式，可以计算出实现该目标的最优输入。对于优化，采用双曲正切函数的 3 段和 5 段分段线性形式，对以双曲正切为激活函数的训练好模型进行混合整数线性规划近似。

在本案例研究中，随着神经元数量增加，问题规模也随之扩大，因此，采用精确双曲正切函数构建的优化问题极具挑战性，难以求解（Schweidtmann & Mitsos，2019）。在 GAMS 中，使用 CPLEX 求解器求解嵌入分段线性神经网络的相应混合整数线性规划问题。随后，将最优输入集值代入由（Alhajri 等人，2008年）得到的数学单元模型，以获得最优输出，并将这些输出表示为第一性原理模型输出（y_{FP}）。在最后一步，如果人工神经网络代理模型的最优解可行，则通过计算公式18中给出的平方误差和（SSE），比较采用人工神经网络模型优化（y^*）计算得到的最优输出，以及与分段线性优化所得最优输入集相对应的第一性原理模型结果。考虑次优训练的影响时，这一比较准则十分重要。计算流程见图5。

$$SSE = \sum_{i=1}^{N}\sum_{j=1}^{6}\left(y_{FP,ij} - y_{i,j}^{*}\right)^{2} \tag{18}$$

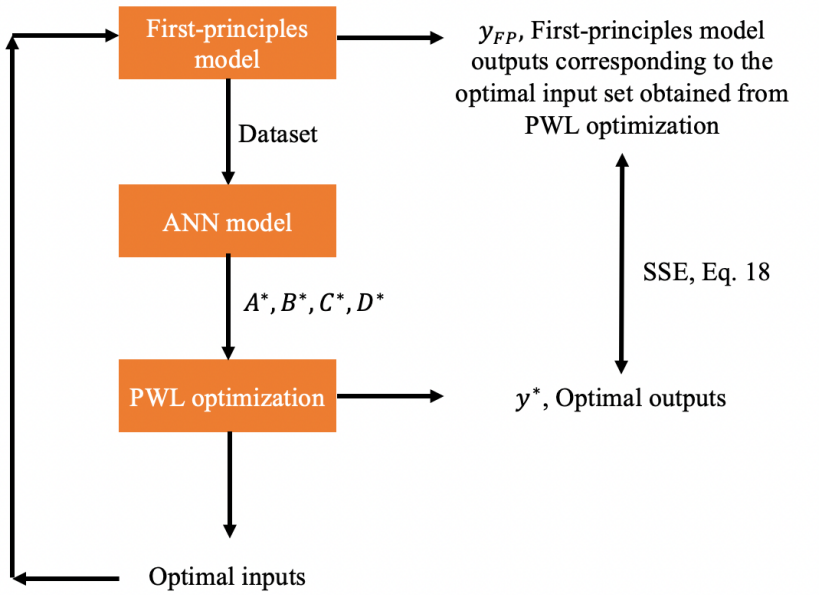

图5。 比较方案。

对于隐藏层中不同数量的神经元，分别计算物理信息引导和非物理信息引导两种情况下可行运行的平均平方误差和，并列于表7。

对于采用双曲正切作为激活函数、采用 Adam 优化器训练的模型，主要有以下三点结论：

（i）在神经元数量相同的情况下，物理信息驱动的模型结果更接近全局解。尽管训练和测试的均方误差值极为接近，优化结果仍可能发生变化，这表明稳健且一致的训练甚至可能不足以实现准确的代理优化。研究发现，物理信息驱动的训练能够实现更可靠的优化。

（ii）使用双曲正切的五段形式时，人工神经网络代理模型优化输出与对应于相同最优输入集的第一性原理模型输出更为接近

。这表明，与其使用精确的双曲正切来构建优化问题并将其作为非线性规划（NLP）问题求解（这尤其是在神经元数量较多、问题规模较大时颇具挑战性），不如利用凸性，将混合整数线性规划公式扩展为包含更多断点和分段的形式。

(iii) 将代理优化得到的最优输入集用于第一性原理数学模型时，可能会出现不可行的情况；当神经网络采用无物理信息目标函数进行训练时，这种情况更有可能发生。这表明，尽管 PI- 和 PI+ 情况下优化得到的解相似，但 PI- 训练和优化仍可能得到物理上无法实现的输入集。另一个结论是，物理信息目标函数可能克服次优训练的影响。

3.3.2. 多目标训练和 ReLU 激活函数

本节对 ReLU 的标准用法与新提出的双曲正切近似进行比较。如前所述，在此类代理优化研究中，使用 ReLU 激活函数是最先进的方法。本节采用隐藏层含有 10 个神经元的网络进行训练，其中使用 ReLU 作为激活函数，输出层则使用恒等函数。 表6中列出的均方误差值表明，使用 ReLU 时，物理信息驱动情形下的训练和测试性能优于无物理信息情形。

reluMIP 工具箱可用于嵌入神经网络模型的代理优化（Lueg 等，2021）。该工具箱利用 ReLU 的精确分段特性，将优化问题表述为混合整数线性规划。不过，ReLU 激活函数是唯一可选项。使用 Gurobi 作为混合整数线性规划求解器，可计算出使原油流量最小化的最优输入。随后，将最优输入传递给第一性原理模型，以获得

与最优输入集对应的输出。 表7中的平均平方误差和表明，物理信息训练带来的偏差要小得多，因此优化结果也更加准确。由此可见，物理信息训练不仅有益于双曲正切激活函数，也有益于 ReLU；对于 ReLU，还可以使用公开可用的开源工具箱进行优化。

3.3.3. 约束训练

人工神经网络模型也在 GAMS 中进行训练，其中可对 PI+情形进行约束训练，称为物理约束训练。对于 PI- 和 PI+ 情形，隐藏层均包含 10 个神经元，并使用双曲正切作为激活函数。由于传统的 TensorFlow 训练算法通常无法求解约束非线性规划问题，因此采用 CONOPT 作为局部训练求解器。

PI- 情形根据公式2进行训练，其中不包含物理项或约束；而 PI+情形则按以下两个步骤进行训练：

a) 首先，训练中不包含物理项或约束，如公式5a所示，这对应于热启动。这一部分与无物理信息训练相同。b) 第二步中，将均方误差和物理项约束 Z 与 U 设为上界，如公式5b所示。根据均方误差和物理约束重新计算权重和偏置参数。

表6显示，与物理信息不驱动的情况相比，物理信息驱动情况下的训练和测试样本在均方误差方面表现更好。这一趋势不同于其他采用隐藏层双曲正切函数训练的神经网络的训练和测试性能，这是由于热启动和约束训练的优势。第一步的热启动

为物理信息训练部分提供了更好的初始化。在第二部分中，根据物理约束调节权重和偏置项，从而提高训练和测试性能。当为物理项定义上界时，训练和测试性能可能会下降，导致欠拟合。通常，这是多目标优化中的典型问题。为避免这一问题，必须在第二部分中同时为物理项和均方误差引入约束。确定上界是约束训练中的另一个问题，因为具有严格上界的训练问题可能会收敛到不可行解。另一方面，当放宽上界时，约束训练的影响可能微乎其微。必须通过反复试验确定上界，直到获得令人满意的训练性能。

优化过程与第 3.3.1 节相同，并采用三段近似来近似双曲正切。通过约束训练实现的物理信息驱动情形能够获得更可靠的优化结果，与全局解的偏差更小；物理信息双目标训练同样如此。

作为训练算法，CONOPT 和 Adam 优化器分别采用广义约化梯度和扩展版随机梯度下降。与 Adam 优化器相比，使用 CONOPT 可获得更高的训练准确率和测试准确率，但采用 CONOPT 训练模型时，最优结果会偏离第一性原理模型的解，这可能是由训练算法和实例之间的差异导致的。尽管如此，关键在于，无论采用何种训练算法，以及物理信息项实现方式（多目标或约束）如何，物理信息训练神经网络都能为代理优化带来更高的准确率和精度。

4. 结论

神经网络的物理信息训练是克服基于物理的模型的复杂性、改善数据驱动模型黑箱特性的一种有前景的方法。本文针对三个不同的案例研究——简单的混合过程、工业原油蒸馏塔和原油装置——研究了与非物理信息训练相比，物理信息训练对过程优化的影响。对于前两个案例研究，质量平衡消除了数据中噪声的影响，因此测试性能有所提升。采用物理信息训练神经网络时，以精确的双曲正切函数构建并由非凸全局求解器求解的过程优化问题会收敛至全局最优解。采用双曲正切激活函数五段近似的混合整数线性规划公式所得结果，与采用精确双曲正切函数的公式所得结果更为接近，显著体现了所提方法用于过程优化的泛化能力。事实上，使用物理信息分段线性神经网络进行代理优化，几乎每次都能获得全局解。此外，由于其凸性，混合整数线性规划公式所需的 CPU 时间大幅减少。

对于第三个案例研究——原油装置，测试了基于求解器、非线性激活函数和分段线性近似的不同训练与优化方法。对于单隐藏层以及包含 10、15、20 个神经元的网络，使用 Adam 优化器进行训练，并将双曲正切作为非线性激活函数。结果表明，对于不同的神经元数量，物理信息案例能够得到更可靠的优化结果。与三段近似相比，双曲正切的五段近似所得结果更接近全局最优解。将隐藏层的激活函数设为 ReLU 时，结果也相同。

最后，在训练问题中将物理项作为约束或次要目标，可提升优化效果，并消除对数据质量和数量的依赖所带来的挑战，而这正是机器学习应用中的一个问题。此外，在

除了使用 ReLU 等天然具有分段线性特性的激活函数外，双曲正切也可以作为一个不错的候选函数，用于构建嵌入物理信息神经网络的混合整数线性规划优化问题。因此，可以增加双曲正切近似中的分段数量，从而获得更接近使用双曲正切精确形式进行优化的结果。最后，使用物理信息驱动的分段线性网络进行过程优化，可显著缩短计算时间，这对于深度神经网络以及实时优化与控制而言，可能是一项优势。

致谢

我们衷心感谢 TÜPRAŞ 炼油厂和 TÜPRAŞ 研发部门分享其实际过程数据。

参考文献

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghe mawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Zheng, X. (2016). *TensorFlow*：大规模机器学习系统。OSDI。

Alhajri, I., Elkamel, A., Albahri, T., & Douglas, P. L. (2008年). 一种用于炼油厂规划和优化的非线性规划模型，采用严格的过程模型和产品质量规格。《国际石油、天然气与煤炭技术杂志》，*1*(3), 283–307. https://doi.org/10.1504/IJOGCT.2008.019846

Ceccon, F., Jalving, J., Haddad, J., Thebelt, A., Tsay, C., Laird, C. D., & Misener, R. (2022).*OMLT*：优化与机器学习工具包。http://arxiv.org/abs/2202.02414 Chollet, F., & et al. (2015). *Keras*。https://keras.io

D'Ambrosio, C., Lodi, A., & Martello, S. (2010). 混合整数线性规划模型中两个变量函数的分段线性近似。《运筹学快报》，*38*(1), 39–46. https://doi.org/10.1016/j.orl.2009.09.005

Di Martino, M., Avraamidou, S., & Pistikopoulos, E. N. (2022). 面向反渗透海水淡化厂的基于神经网络的超结构优化方法。《膜》，*12*(2). https://doi.org/10.3390/membranes12020199

Grimstad, B., & Andersson, H. (2019). ReLU网络作为混合整数线性规划中的代理模型。《计算机与化学工程》，*131*。https://doi.org/10.1016/j.compchemeng.2019.106580

Haghighat, E., & Juanes, R.（2021年）。SciANN：一个用于科学计算和基于人工神经网络的物理信息深度学习的 Keras/TensorFlow 封装器。《计算机方法在应用力学与工程中的应用》，*373*。https://doi.org/10.1016/j.cma.2020.113552

Karuppiah, R., & Grossmann, I. E.（2006）。化学过程集成水系统综合的全局优化。《计算机与化学工程》，*30*(4)，650–673。
https://doi.org/10.1016/j.compchemeng.2005.11.005

Katz, J., Pappas, I., Avraamidou, S., & Pistikopoulos, E. N. (2020a). 显式模型预测控制与基于 ReLU 的神经网络的集成。*IFAC-PapersOnLine*，*53*(2), 11350–11355.
https://doi.org/10.1016/j.ifacol.2020.12.544

Katz, J., Pappas, I., Avraamidou, S., & Pistikopoulos, E. N. (2020b). 深度学习模型与多参数规划的集成。《计算机与化学工程》，*136*.
https://doi.org/10.1016/j.compchemeng.2020.106801

Kingma, D. P., & Ba, J. (2014年). *Adam*：一种随机优化方法.
http://arxiv.org/abs/1412.6980

Lueg, L., Grimstad, B., Mitsos, A., & Schweidtmann, A. M. (2021年). *reluMIP*：用于 *ReLU* 神经网络混合整数线性规划优化的开源工具.
https://github.com/ChemEngAI/ReLU_ANN_MILP

Masi, F., Stefanou, I., Vannucci, P., & Maffi-Berthier, V.（2021年）。用于本构建模的基于热力学的人工神经网络。固体力学与物理学杂志，*147*。
https://doi.org/10.1016/j.jmps.2020.104277

Merkelbach, K., Schweidtmann, A. M., Müller, Y., Schwoebel, P., Mhamdi, A., Mitsos, A., Schuppert, A., Mrziglod, T., & Schneckener, S.（2022）。HybridML：用于混合建模的开源平台。《计算机与化学工程》，*160*。
https://doi.org/10.1016/j.compchemeng.2022.107736

Nellikkath, R.，& Chatzivasileiadis, S.（2021a）。用于最小化直流最优潮流中最坏情况违约的物理信息神经网络。*2021*年 *IEEE* 国际通信、控制与智能电网计算技术会议，*SmartGridComm 2021*，419–424。
https://doi.org/10.1109/SmartGridComm51999.2021.9632308

Nellikkath, R.， & Chatzivasileiadis, S.（2021b）。用于交流最优潮流的物理信息神经网络。http://arxiv.org/abs/2110.02672

Pun, G. P. P., Batra, R., Ramprasad, R., & Mishin, Y.（2019年）。用于材料原子尺度建模的物理信息人工神经网络。《自然·通讯》，*10*(1)。
https://doi.org/10.1038/s41467-019-10343-5

Schweidtmann, A. M., & Mitsos, A.（2019年）。嵌入人工神经网络的确定性全局优化。《优化理论与应用杂志》，*180*(3)，925–948。
https://doi.org/10.1007/s10957-018-1396-0

Sildir, H., & Aydin, E.（2022）。一种基于混合整数线性规划的人工神经网络训练与特征选择方法，采用分段线性近似。《化学工程科学》，*249*。
https://doi.org/10.1016/j.ces.2021.117273

Sridhar, S., Linderoth, J., & Luedtke, J.（2013）。含指示变量的分段线性函数的局部理想公式。《运筹学快报》，*41*(6)，627–632。
https://doi.org/10.1016/j.orl.2013.08.010

Stiasny, J., Chevalier, S., & Chatzivasileiadis, S. (2021年). 无数据学习：用于快速时域仿真的物理信息神经网络；无数据学习：用于快速时域仿真的物理信息神经网络。*2021年 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*.
https://doi.org/10.1109/SMARTGRIDCOMM51999.2021.963

Thebelt, A., Wiebe, J., Kronqvist, J., Tsay, C., & Misener, R. (2022). 从化学工程数据集中最大限度地获取信息：机器学习的应用。《化学工程科学》，*252*。
https://doi.org/10.1016/j.ces.2022.117469

Tsay, C., Kronqvist, J., Thebelt, A., & Misener, R.（2021年）。基于划分的训练后*ReLU*神经网络混合整数优化表述。http://arxiv.org/abs/2102.04373

Vielma, J. P.（2015）。混合整数线性规划建模技术。载于*SIAM*评论（第57卷，第1期，第3–57页）。工业与应用数学学会出版物。https://doi.org/10.1137/130915303

Yang, S. 和 Bequette, B. W.（2021年）。使用输入凸神经网络的基于优化的控制。《计算机与化学工程》，144。https://doi.org/10.1016/j.compchemeng.2020.107143