

多层感知机的解析表达在过程系统优化中的应用

张子轩^{1,2}, 宋晓维^{1,2}, 曾玉娇¹, 李杰³, 朱闽¹, 聂亚玲¹, 陆冬云¹, 王利民^{1,2}, 肖焱^{1*}

(¹中国科学院过程工程研究所, 北京市 100190; ²中国科学院大学化学工程学院, 北京市 100049; ³曼彻斯特大学工程学院化学工程系过程集成中心, 曼彻斯特 M13 9PL)

摘要: 化工过程优化问题常常涉及高度非线性函数或非解析形式的机器学习模型, 使得传统优化方法难以求解。本文使用线性整流单元 (ReLU) 激活的多层感知机 (MLP) 作为代理模型, 利用大 M 法重新表达其解析式, 以解决此类复杂优化问题。将该模型应用于部分近似 (混合整数二次约束规划) 和完全近似 (混合整数线性规划) 的换热网络优化问题, 所获得的解与实际最优解的误差仅为 0.16‰; 在数据驱动的乙酸乙酯生产工况全局优化问题中, 收敛速度相对于非解析形式模型提高 78.35%。该结果验证了使用多层感知机解析表达进行优化的有效性和准确性, 为过程系统优化提供了一种新的有效工具。

关键词: 非线性函数; 机器学习解析表达; 混合整数; 换热网络; 生产工况; 全局优化

中图分类号: TQ 019 **文献标识码:** A

Application of multi-layer perceptron with analytic formulation in optimizing process systems

Zhang Zixuan^{1,2}, Song Xiaowei^{1,2}, Zeng Yujiao¹, Li Jie³, Zhu Min¹, Nie Yaling¹, Lu Dongyun¹, Wang Limin^{1,2}, Xiao Xin^{1*}

(¹Institute of Processing Engineering, Chinese Academy of Science, Beijing 100190, China; ²School of Chemical Engineering, University of Chinese Academy of Science, Beijing 100049, China; ³Centre for Process Integration, Department of Chemical Engineering, School of Engineering, The University of Manchester, Manchester M13 9PL, United Kingdom)

Abstract: Chemical process optimization problems often involve highly nonlinear functions or non-analyzed forms of machine learning models, making traditional optimization methods difficult to solve. This paper uses a multi-layer perceptron (MLP) activated by linear rectification units (ReLU) as a surrogate model, and reformulates its analytic formulation with the Bin-M method to address these complex optimization challenges. Applying this model to partially approximated (mixed integer quadratic constrained programming, MIQCP) and fully approximated (mixed integer linear programming, MILP) heat exchanger network optimization problems, the error between the obtained solution and the actual optimal solution is only 0.16 ‰; In the case of data-driven global optimization for ethyl acetate operating conditions, the convergence speed is increased by 78.35% compared to non-analyzed model. These results validate the effectiveness and accuracy of using analytic formulation of multi-layer perceptron for optimization, providing a new and effective tool for process system optimization.

Key words: Nonlinear functions; Machine learning models with analytic formulation; Mixed integer; Heat exchange network; Production conditions; Global optimization

第一作者: 张子轩 (2001—), 男, 硕士研究生, 联系方式: zhangzixuan23@mails.ucas.ac.cn。

通讯作者: 肖焱, 男, 博士, 研究员, 联系方式: xxiao@ipe.ac.cn。

引言

使用数学模型来描述过程单元与系统是被广泛应用并持续发展的先进技术^[3]。传统的建模优化框架主要依赖于模拟,即使用模拟软件(如 Aspen Plus, PRO-II 等)映射输入输出实现目标优化。对于大规模复杂系统,建立高精度的机理模型既困难又昂贵,因此建立代理模型是一种有效的替代方式。当前最常见的代理模型是数据驱动的机器学习模型。机器学习模型在使用过程中通常被视为一个非解析形式的模型,主要通过遗传算法、粒子群算法等元启发式算法进行求解。元启发式算法的优点是能够处理各种形式的优化问题^[4],能够在大规模搜索空间中找到质量比较好的可行解,但不能保证可行解的最优性。

相较于非解析形式表达的机器学习模型,有明确解析式的优化模型可用于成熟的商业求解器,收敛性和最优性有一定理论保障。因此,基于机器学习模型构建或近似的非线性系统,可使用其解析表达进行优化。

Fischetti 等^[5]最早提出了具有线性整流单元(ReLU)激活的多层感知机模型的大 M 表示。Tsay 等^[6]则对于 ReLU 函数提出了具有更紧界的近似凸包络松弛。Schweidtmann 等^[7]设计并研究了几种同时考虑输入和输出限制的边界收紧程序,可以显著缩短求解时间。对于决策树模型,Maragno 等^[8]提出了大 M 法混合整数线性模型,Biggs 等^[9]和 Mišić^[10]则提出了随机森林混合整数线性方法。Bertsimas 等^[11]将线性模型决策树混合整数形式引入非线性模型的全局优化,讨论了不可行问题的松弛与模型的鲁棒性^[12]。Ammari 等^[13]则研究了几个将线性模型决策树混合整数形式应用于实际工程优化的案例。这些研究仅从松弛问题边界的松紧与全局性进行研究,并未考虑实际模型中近似程度对求解效率的影响,也未对比元启发算法与确定性算法之间的优劣。

鉴于大型复杂系统在于过程系统工程领域广泛存在,建立代理模型求解是值得探索的重要方向。因此,有必要进一步研究拓展机器学习解析表达式在优化任务上的应用。本文以换热网络优化问题为对象,使用 ReLU 函数激活的多层感知机模型进行混合整数近似,测试不同近似程度对模型求解的难易程度与精度。以乙酸乙酯生产系统的参数优化问题作为研究对象,对比了使用确定性算法与启发式算法的求解效率与解的质量。本文从模型近似程度和求解算法两个方面,对使用多层感知机解析形式表达的真实优化问题进行了测试分析,旨在探索指导实际工程优化问题的创新路径。

1. ReLU 函数激活的多层感知机解析表达标准形式

前馈神经网络是每一种无反馈,信号从输入层向输出层单向传播的神经网络。其中,常为过程系统预测任务所用的全连接前馈神经网络(多层感知机)属功能较为强大的机器学习模型。一个包含足够多隐含层神经元的多层感知机,能以任意精度逼近任意预定的连续函数^[14]。

多层感知机由输入层、输出层以及隐藏层组成,特征经过隐藏层后会经过激活函数引入非线性,常用的激活函数有 Sigmoid 函数、ReLU 函数等。使用最广泛的激活函数是 ReLU 函数,且有大 M 法表示的混合整数公式^[5,15],如式(1)所示。ReLU 函数还有多选法^[16]和分段线性函数^[17]等表示方法,以及光滑非线性激活函数的显式表达形式^[18]。

$$u = \max\{0, x\} \Leftrightarrow \begin{cases} x \leq u \leq x + M(1 - \gamma) \\ 0 \leq u \leq M\gamma \\ \gamma \in \{0, 1\} \end{cases} \quad (1)$$

式中: M 为一个很大的正数; γ 为二元松弛变量; x 为激活前函数值; u 为激活后的函数值。

考虑具有 n 个隐藏层的多层感知机,其显式表达可以用如下一组约束表示:

$$\begin{aligned}
y_{MLP} &= W_{out}x_{out}^{h_n} + b_{out} \\
x_{out}^{h_i} &= \Phi(x_{in}^{h_i}) \quad i = 1, \dots, n \\
x_{in}^{h_{i+1}} &= W_{i,i+1}x_{out}^{h_i} + b_{i,i+1} \quad i = 1, \dots, n-1 \\
x_{in}^{h_1} &= W_{in}x + b_{in}
\end{aligned} \tag{2}$$

式中： y_{MLP} 为多层感知机的预测值； W_{in} 、 W_{out} 和 $W_{i,i+1}$ 分别为输入层到第1个隐藏层、第n个隐藏层到输出层、第*i*个隐藏层到第*i+1*个隐藏层的权重； b_{in} 、 b_{out} 和 $b_{i,i+1}$ 分别为输入层到第1个隐藏层、第n个隐藏层到输出层、第*i*个隐藏层到第*i+1*个隐藏层的偏置； x 为输入特征； $x_{in}^{h_i}$ 和 $x_{out}^{h_i}$ 分别为第*i*个隐藏层激活前后的值； $\Phi(\cdot)$ 为激活函数。

若原始约束为不等式约束，可增加公式(3)：

$$y_{MLP} \leq 0 \tag{3}$$

动模型自身的解析化。多层感知机模型相比于其他机器学习模型的巨大优势在于可以具有高维的输入与输出，特别适用于复杂过程系统建模。因此本文选择具有 ReLU 函数激活的多层感知机作为机器学习模型，考察其解析表达在换热网络优化及数据驱动生产工况优化上的应用。

2.1 换热网络优化的混合整数线性近似

换热网络优化问题是过程系统工程中的经典优化问题。Yee 等^[19]提出的分级超结构模型已经被广泛接受并应用于工业设计优化当中。然而，由于基于超结构设计的模型是 MINLP 模型，大规模问题求解十分困难，无论是求解速度还是全局最优性都无法保证。因此，本文尝试使用多层感知机对换热网络中的非线性函数做混合整数近似，以提高求解速度，并分析解的质量。

本文选取的例子来源于 Yee 等^[19]的四流股无分流超结构模型，其流股数据如表 1 所示。

表 1 Yee 等四流股无分流案例数据

Tab. 1 Data of four streams without splitting form Yee's work

Stream number	Temperature, K		Average heat capacity flow rate, kW·K ⁻¹	Cost, \$·kW ⁻¹ ·yr ⁻¹
	inlet	outlet		
H1	443	333	30	
H2	423	303	15	
C1	293	408	20	
C2	353	413	40	
Steam	450	450		80
Cooling Water	293	313		20

$U = 0.8$ (kW·m⁻²·K⁻¹) for all matches except ones involving steam

$U = 1.2$ (kW·m⁻²·K⁻¹) for all matches involving stream

Annual cost = $1000 \times [\text{Area}(\text{m}^2)]^{0.6}$ (\$·yr⁻¹) for all exchangers except heaters

Annual cost = $1200 \times [\text{Area}(\text{m}^2)]^{0.6}$ (\$·yr⁻¹) for heaters

2. 案例分析

机器学习解析表达主要有两种应用场景，第一类场景是对于已知的有解析形式的非线性模型进行混合整数线性近似；另一类场景则是对于数据驱

确定性优化求解器使用 Gurobi Optimizer 11.0.0 版本，内部版本为 v11.0.0rc2 (win64-Windows 11+0 (22631.2))。CPU 型号：AMD Ryzen 7 7840H，Radeon 780M Graphics，instruction set [SSE2|AVX|AVX2|AVX512]。线程数：8 个物理内

核, 16 个逻辑处理器, 使用 16 个线程。

2.1.1 换热网络 MINLP 模型

无分流换热网络模型的约束条件如下:

1) 总流股热量平衡

$$\begin{aligned} (T_i^{in} - T_i^{out})F_i &= \sum_{k \in S} \sum_{j \in C} Q_{i,j,k} + Q_i^{cu} \quad i \in H \\ (T_j^{out} - T_j^{in})F_j &= \sum_{k \in S} \sum_{i \in H} Q_{i,j,k} + Q_j^{hu} \quad j \in C \end{aligned} \quad (4)$$

式中: 下标 j, i 为冷热流股索引, k 为超结构级索引; 集合 C, H 为冷热流股集合, S 为超结构级集合; 上标 in, out 表示进出口, cu, hu 表示冷热公用工程换热器; 参数 F ($\text{kW} \cdot \text{K}^{-1}$) 为热容流率, T (K) 为温度; 变量 Q (kW) 为换热量。

2) 超结构级热量平衡

$$\begin{aligned} (t_{i,k} - t_{i,k+1})F_i &= \sum_{j \in C} Q_{i,j,k} \quad k \in S, i \in H \\ (t_{j,k} - t_{j,k+1})F_j &= \sum_{i \in H} Q_{i,j,k} \quad k \in S, j \in C \end{aligned} \quad (5)$$

式中: 变量 t (K) 为超结构级的边界温度。

3) 超结构入口温度

$$\begin{aligned} T_i^{in} &= t_{i,1} \quad i \in H \\ T_j^{in} &= t_{j,N+1} \quad j \in C \end{aligned} \quad (6)$$

式中: 参数 N 为超结构级数。

4) 温度可行性

$$\begin{aligned} t_{i,k} &\geq t_{i,k+1} \quad k \in S, i \in H \\ t_{j,k} &\geq t_{j,k+1} \quad k \in S, j \in C \\ t_{i,N+1} &\geq T_i^{out} \quad i \in H \\ T_j^{out} &\geq t_{j,1} \quad j \in C \end{aligned} \quad (7)$$

5) 公用工程负荷

$$\begin{aligned} (t_{i,N+1} - T_i^{out})F_i &= Q_i^{cu} \quad i \in H \\ (T_j^{out} - t_{j,1})F_j &= Q_j^{hu} \quad j \in C \end{aligned} \quad (8)$$

6) 负荷逻辑约束

$$\begin{aligned} Q_{i,j,k} - \Omega z_{i,j,k} &\leq 0 \quad k \in S, i \in H, j \in C \\ Q_i^{cu} - \Omega z_i^{cu} &\leq 0 \quad i \in H \\ Q_j^{hu} - \Omega z_j^{hu} &\leq 0 \quad j \in C \\ z_{i,j,k}, z_i^{cu}, z_j^{hu} &\in \{0, 1\} \quad k \in S, i \in H, j \in C \end{aligned} \quad (9)$$

式中: z 为 0-1 逻辑变量, 代表流股之间是否存在换热; Ω 为大 M 法的常数, 为了获得较紧的上界,

可令 $\Omega = \max\{(T_i^{in} - T_i^{out})F_i, (T_j^{out} - T_j^{in})F_j : i \in H, j \in C\}$ 。

7) 换热温差逻辑约束

$$\begin{aligned} \Delta t_{i,j,k} &\leq t_{i,k} - t_{j,k} + \Gamma(1 - z_{i,j,k}) \quad k \in S, i \in H, j \in C \\ \Delta t_{i,j,k+1} &\leq t_{i,k+1} - t_{j,k+1} + \Gamma(1 - z_{i,j,k}) \quad k \in S, i \in H, j \in C \\ \Delta t_i^{cu} &\leq t_{i,N+1} - T_{cu}^{out} + \Gamma(1 - z_i^{cu}) \quad i \in H \\ \Delta t_j^{hu} &\leq T_{hu}^{out} - t_{j,1} + \Gamma(1 - z_j^{hu}) \quad j \in C \end{aligned} \quad (10)$$

式中: Δt (K) 为流股之间的换热温差; T_{cu}^{out} (K) 和

T_{hu}^{out} (K) 分别为冷热公用工程的出口温度; Γ 大 M

法的常数, 为了获得较紧的上界, 可令

$\Gamma = \max\{(T_i^{in} - T_j^{in}), (T_i^{out} - T_j^{out}) : i \in H, j \in C\}$ 。

8) 最小换热温差约束

$$\Delta t \leq \Delta T_{\min} \quad (11)$$

式中: ΔT_{\min} (K) 为最小换热温差。

无分流换热网络的目标函数如下:

$$\begin{aligned} obj = \min & \sum_{i \in H} \text{Cost}^{cu} Q_i^{cu} + \sum_{j \in C} \text{Cost}^{hu} Q_j^{hu} \\ & - \sum_{i \in H} \text{Cost}^{fix} z_i^{cu} + \sum_{j \in C} \text{Cost}^{fix} z_j^{hu} + \sum_{k \in S} \sum_{i \in H} \sum_{j \in C} \text{Cost}^{fix} z_{i,j,k} \\ & - \sum_{i \in H} 1000 \times (\text{area}_i^{cu})^{0.6} + \sum_{j \in C} 1200 \times (\text{area}_j^{hu})^{0.6} + \sum_{k \in S} \sum_{i \in H} \sum_{j \in C} 1000 \times (\text{area}_{i,j,k})^{0.6} \end{aligned} \quad (12)$$

式中: 参数 Cost^{cu} ($\text{\$} \cdot \text{kW}^{-1} \cdot \text{yr}^{-1}$) 和 Cost^{hu} 分别为冷热

公用工程的价格, Cost^{fix} ($\text{\$} \cdot \text{yr}^{-1}$) 为换热器固定投资

费用; 变量 area (m^2) 为换热面积。

其中:

$$\text{area}_{i,j,k} = \frac{Q_{i,j,k}}{U \cdot \text{MDT}_{i,j,k}} \quad (13)$$

式中: 参数 U ($\text{kW} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$) 为换热系数; 变量 MDT (K) 为平均换热温差。

Yee 等^[19]使用 Chen^[20]所提出的平均换热温差来近似对数平均换热温差:

$$MDT_{i,j,k} = \left(\frac{\Delta t_{i,j,k}^2 \Delta t_{i,j,k+1} + \Delta t_{i,j,k} \Delta t_{i,j,k+1}^2}{2} \right)^{\frac{1}{3}} \quad (14)$$

本文直接使用对数平均换热温差的原始公式:

$$LMTD_{i,j,k} = \frac{\Delta t_{i,j,k} - \Delta t_{i,j,k+1}}{\ln(\Delta t_{i,j,k}) - \ln(\Delta t_{i,j,k+1})} \quad (15)$$

本文使用多层感知机对对数平均温差公式(15)以及目标函数(12)中幂函数的进行近似。多层感知机的结构会影响最终解的质量和求解速度。为此,本文进行多层感知机结构的超参数调优。在本文中,主要考虑了网络层数与隐藏层神经元个数两个超参数,采用粒子群算法进行了超参数寻优。同时,为了使近似后的优化模型不太复杂,需要限制超参数的最大值,为此我们还限制模型隐藏层总层数不超过3层,每层神经元数目不超过100个。

2.1.2 使用多层感知机部分近似的 MIQCP 模型

2.1.1 中的 MINLP 模型非线性主要源于式(15),因此首先考虑使用多层感知机对式(15)进行混合整数线性近似。所建立的多层感知机共1个隐藏层,每个隐藏层含有100个神经元,隐藏层使用 ReLU 函数激活。采样范围为 0-300 K,测试集均方误差为 0.0119。式(15)可以使用如式(16)所示的一组约束表示。

$$\begin{aligned} LMTD_{i,j,k} &= W_{out} x_{i,j,k}^{out} + b_{out} \\ x_{i,j,k}^{in} &= W_{in} \begin{bmatrix} \Delta t_{i,j,k} \\ \Delta t_{i,j,k+1} \end{bmatrix} + b_{in} \\ x_{i,j,k}^{in} &\leq x_{i,j,k}^{out} \leq x_{i,j,k}^{in} + M(1 - \gamma_{i,j,k}) \\ 0 &\leq x_{i,j,k}^{out} \leq M \gamma_{i,j,k} \\ \gamma_{i,j,k} &\in \{0,1\} \end{aligned} \quad (16)$$

式中:输入特征为 $[\Delta t_{i,j,k} \quad \Delta t_{i,j,k+1}]^T$, 预测值为

$LMTD_{i,j,k}$; 参数 $W_{in}, b_{in}, W_{out}, b_{out}$ 分别为 100×2 、 100×1 、 1×100 、 1×1 的矩阵,分别代表输入层权重和偏置、输出层权重和偏置, M 为大 M 法常数; 变量 $x_{i,j,k}^{in}$ 和 $x_{i,j,k}^{out}$ 分别为特征经过激活层前后的值,

均为 100×1 的矩阵, $\gamma_{i,j,k}$ 为大 M 法的二元逻辑辅助变量。

此时,式(12)中仅剩 $(area)^{0.6}$ 一项非线性项,可以引入松弛变量使其转化为二次约束,则该模型转化为非凸 MIQCP 模型。

$$\begin{aligned} r &= (area)^{0.6} \Leftrightarrow r^5 = area^3 \\ &\Rightarrow \begin{cases} r_1 = area^2 & r_4 = r_3^2 \\ r_2 = r_1 area & r_5 = r r_4 \\ r_3 = r^2 & r_2 = r_5 \end{cases} \end{aligned} \quad (17)$$

式中: r, r_i 为松弛变量。

2.1.3 使用多层感知机完全近似的 MILP 模型

本节考虑对 $\left(\frac{Q_{i,j,k}}{LMTD_{i,j,k}} \right)^{0.6}$ 一项进行混合整数

线性近似。对 $Q_{i,j,k}$ 在 0-3500 kW 采样,对 $LMTD_{i,j,k}$ 在 0-180 K 采样,采样方法为拉丁超立方采样。由于特征尺度不同,首先对特征进行标准化。所建立的多层感知机共2个隐藏层,每个隐藏层含有20个神经元,隐藏层使用 ReLU 函数激活,测试集均方误差为 0.334。该非线性项可以用如式(18)所示的一组约束表示。

$$\begin{aligned} (y_{i,j,k} - y_{i,j,k}^{mean}) / (y_{i,j,k}^{var})^{0.5} &= W_{out} x_{i,j,k}^{h_2, out} + b_{out} \\ x_{i,j,k}^{h_1, in} &= W_{in} \begin{bmatrix} (LMTD_{i,j,k} - LMTD_{i,j,k}^{mean}) / (LMTD_{i,j,k}^{var})^{0.5} \\ (Q_{i,j,k} - Q_{i,j,k}^{mean}) / (Q_{i,j,k}^{var})^{0.5} \end{bmatrix} + b_{in} \\ x_{i,j,k}^{h_2, in} &= W_{12} x_{i,j,k}^{h_1, out} + b_{12} \\ x_{i,j,k}^{h_1, in} &\leq x_{i,j,k}^{h_1, out} \leq x_{i,j,k}^{h_1, in} + M(1 - \gamma_{i,j,k}^h) \\ 0 &\leq x_{i,j,k}^{h_1, out} \leq M \gamma_{i,j,k}^h \\ x_{i,j,k}^{h_2, in} &\leq x_{i,j,k}^{h_2, out} \leq x_{i,j,k}^{h_2, in} + M(1 - \gamma_{i,j,k}^{h_2}) \\ 0 &\leq x_{i,j,k}^{h_2, out} \leq M \gamma_{i,j,k}^{h_2} \\ \gamma_{i,j,k}^h, \gamma_{i,j,k}^{h_2} &\in \{0,1\} \end{aligned} \quad (18)$$

式中:输入特征为 $\left[(LMTD_{i,j,k} - LMTD_{i,j,k}^{mean}) / (LMTD_{i,j,k}^{var})^{0.5} \quad (Q_{i,j,k} - Q_{i,j,k}^{mean}) / (Q_{i,j,k}^{var})^{0.5} \right]^T$, 预测值为 $y_{i,j,k}$; 上标 $mean$ 和 var 分别表示特征的平均值和方差,带有该上标的量为参数。

目标函数(12)可用如下公式表示:

$$\begin{aligned}
 obj = \min & \sum_{i \in H} Cost^{cu} Q_i^{cu} + \sum_{j \in C} Cost^{hu} Q_j^{hu} \\
 & + \sum_{i \in H} Cost^{fix} z_i^{cu} + \sum_{j \in C} Cost^{fix} z_j^{hu} + \sum_{k \in S} \sum_{i \in H} \sum_{j \in C} Cost^{fix} z_{i,j,k} \\
 & + \sum_{i \in H} \frac{1000}{0.8^{0.6}} \times y_i^{cu} + \sum_{j \in C} \frac{1200}{1.2^{0.6}} \times y_j^{hu} + \sum_{k \in S} \sum_{i \in H} \sum_{j \in C} \frac{1000}{0.8^{0.6}} \times y_{i,j,k}
 \end{aligned} \quad (19)$$

该模型转化为 MILP 模型。

2.1.4 结果对比

三种模型的规模与求解结果如表 2 所示, 最优解详细信息如表 3 所示。原始文献与三种模型求解得到的网络结构均如图 1 所示。由于使用机器学习模型拟合会有一些误差, 因此根据所得换热器负荷对数据进行校正, 括号内为校正后的真实值。

三种模型下求解结果均优于原始文献结果, 其中 MIQCP 模型在相对较短时间内找到理论最优解, 与原始严格模型最优值十分接近。MILP 模型相对 MINLP 模型在相同时间内求得的对偶间隙更小, 但由于大量二元变量存在, 因此也存在收敛困难的问题。

对比线性化程度低的 MIQCP 中共有 2844 个二元变量, 共有 2^{2844} 个分支需要搜索, 目标函数最优值与校正值相对误差为 0.295%; 而 MILP 模型共有 3484 个二元变量, 即共有 2^{3484} 个分支需要搜索, 目标函数最优值与校正值相对误差为 14.1%。因此线性化程度越高不一定会使模型更易求解, 反而可能导致更大的误差与求解难度。

表 2 换热网络优化模型规模与求解对比

Tab. 2 Comparison of scale and solution of heat exchange network optimization models

	MINLP	MIQCP	MILP
Integer Variables	44	2844	3484
Termination mode	Time limit reached	Optimal solution found	Time limit reached
Termination time/s	9999.02	424.95	10001.17
Max constraint violation	3.1841×10^{-2}	3.1818×10^{-2}	-
Best objective	8.1589×10^{-4}	8.1893×10^{-4}	7.1571×10^{-4}
Best bound	5.6920×10^{-4}	8.1887×10^{-4}	6.7932×10^{-4}
Gap	30.2355%	0.0071%	5.0845%

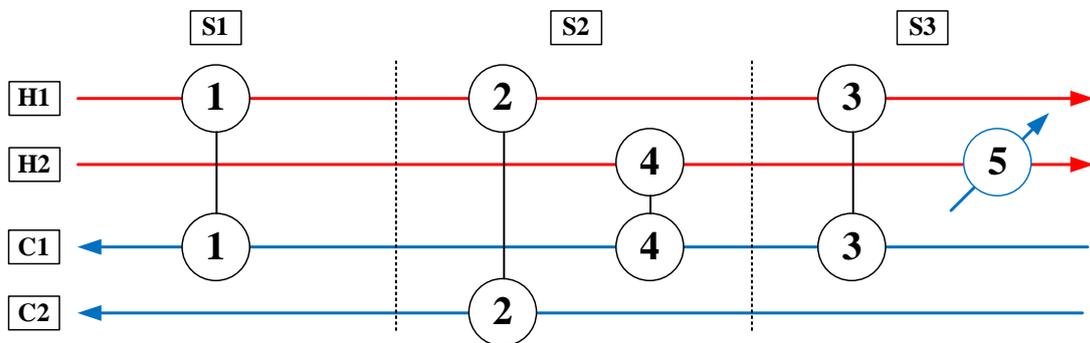


图 1 换热网络结构图

Fig. 1 Structure diagram of heat exchange network

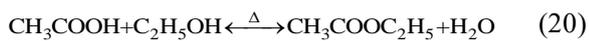
表 3 不同模型下换热网络最优解对比

Tab. 3 Comparison of optimal solutions with different heat exchange networks models

Model	Exchanger	Heat load/kW	Area/m ²	(Area) ^{0.6}	Annual utility cost/\$·yr ⁻¹	Annual capital cost/\$·yr ⁻¹	Total annual cost/\$·yr ⁻¹
Original literature	1	219.6	7.5(7.55)	-			
	2	2400	320.3(327.21)	-		72909	80909
	3	680.4	25.0(25.00)	-	8000	(73857.67)	(81857.67)
	4	1400	171.3(178.08)	-			
	5	400	38.3(38.31)	-			
MINLP	1	218.47	7.43(7.42)	-			
	2	2400	325.19(325.19)	-		73589	81589
	3	681.53	25.08(25.05)	-	8000	(73651.16)	(81651.16)
	4	1400	177.60(177.57)	-			
	5	400	38.31(38.31)	-			
MIQCP	1	217.30	7.39(7.39)	-			
	2	2400	323.13(324.99)	-		73893	81893
	3	682.70	25.10(25.13)	-	8000	(73651.29)	(81651.29)
	4	1400	179.30(181.59)	-			
	5	400	38.31(38.31)	-			
MILP	1	215.35	7.32(7.32)	3.33 (3.30)			
	2	2400	321.20(319.78)	32.13 (31.84)			
	3	684.65	25.21(25.19)	6.90 (6.93)	8000	63571 (73674.66)	71571 (81674.66)
	4	1400	184.77(182.31)	22.35 (22.72)			
	5	400	38.30 (38.31)	8.90 (8.91)			

2.2 数据驱动的生产工况优化

现考虑一个简单的乙酸乙醇制乙酸乙酯的反应器，存在如下反应：



假设系统进料摩尔流量一定，进料水、乙醇和乙酸的摩尔分率及反应温度为可操作变量。使用 Aspen Plus 平推流反应器进行建模，并使用拉丁超立方采样，特征变量为反应进料乙醇、乙酸的摩尔

分率及反应温度，标签变量为反应后乙酸乙酯的质量流量和进料热负荷，具体参数设置如表 4 所示。

采样并进行剔除异常值后得到 9425 个数据点，使用 ReLU 函数激活的多层感知机模型训练，该模型含有 3 个隐藏层，每个隐藏层含有 20 个神经元，训练得到的模型在测试集的相对误差为 0.019。

表 4 模拟参数设置

Tab. 4 Parameters of the simulation

参数	值
进料流量/ $\text{kmol}\cdot\text{h}^{-1}$	388
反应器类型	绝热
反应器直径/m	0.3
反应器长度/m	2
正反应指前因子	1.9×10^8
正反应活化能/ $\text{J}\cdot\text{kmol}^{-1}$	5.95×10^7
逆反应指前因子	5×10^7
逆反应活化能/ $\text{J}\cdot\text{kmol}^{-1}$	5.95×10^7

优化模型如下：

$$\min \sum_{i \in R} \text{cost}_i^r x_i F + \text{cost}^u q - \sum_{j \in P} \text{price}_j m_j$$

$$= 74511.744x_{\text{acetic}} + 124962.404x_{\text{ethanol}} + 0.0029q - 10m_{\text{ethyl}} \quad (21)$$

$$\text{s.t. } ML(x_i, T) \longrightarrow q, m_j$$

$$\sum_{i \in R} x_i = 1$$

$$T \in [0, 200] \quad x_i \in [0, 1]$$

式中：集合 R 为反应物集合， P 为产物集合；上标 r 为反应物， u 为公用工程；下标 acetic 为乙酸， ethanol 为乙醇， ethyl 为乙酸乙酯；参数 cost ($\text{\$}\cdot\text{kmol}^{-1}$, $\text{\$}\cdot\text{kW}^{-1}\cdot\text{hr}^{-1}$) 为花费价格， price ($\text{\$}\cdot\text{ton}^{-1}$) 为售价， F ($\text{kmol}\cdot\text{hr}^{-1}$) 为总流量；变量 x 为摩尔分率， q (kW) 为反应器加热功率， m ($\text{ton}\cdot\text{hr}^{-1}$) 为产物质量， T (K) 为反应器温度；函数 $ML(\cdot)$ 为训练得到的多层感知机模型。

将机器学习解析表达问题转化为混合整数线性规划后，调用 Gurobi 求解器求解；对于非解析表达模型，采用罚函数法进行松弛，并调用粒子群算法求解，求解结果如表 5 所示。显然，对于直接建立数据驱动模型的工况优化问题，使用解析形式表达的模型相对于黑箱模型，在解的质量和收敛时间上均有显著提升。在相对短的时间内，非解析表达模型也能得到质量相对好的可行解，但要想获得全局最优解则需耗费更多时间。

表 5 乙酸乙酯系统优化求解结果对比

Tab. 5 Comparison of optimization results for ethyl acetate

	system			
	解析表 达	非解析表达		
种群数量	-	500	2000	
收敛时间/s	5.76	10.98	26.60	
x_{ethanol}	0.44	0.44	0.44	
x_{acetic}	0.56	0.56	0.56	
T /K	30.7	30.8	30.7	
最优解	q /kW	124.62	126.12	124.62
	m_{ethyl}	10.86	10.85	10.86
	/ton·hr ⁻¹	-	-	-
	最优值/CNY·hr ⁻¹	11838.33	11835.79	11838.33

3. 结论

本文以换热网络和生产工况优化为例，探讨了 ReLU 函数激活多层感知机的解析形式在过程系统优化中的应用。通过 ReLU 函数激活的多层感知机模型拟合非线性函数或系统，通过大 M 松弛法解析表达模型，案例分析表明，使用这种解析表达的优化模型有效。在换热网络优化案例中，部分近似的 MIQCP 模型展现出最佳性能，既保证了较小的误差，又在有限时间内收敛至全局最优解。对于数据驱动模型的直接解析优化，解析形式的模型配合确定性求解器性能明显优于非解析模型配合启发式优化算法。在乙酸乙酯生产案例中，与非解析模型相比，解析模型收敛到相同最优值时，解析模型收敛速度提高 78.35%。同时发现，对于 MINLP 问题中的非线性函数进行近似时，由于引入了大量二元变量，线性化程度越高不一定会使模型更易求解，反而可能导致更大的误差与求解难度。因此，对于具有特定形式的使用机器学习解析表达建立的优化模型，其高效求解算法仍有待开发。

参考文献(References):

- [1] RIOS L M, SAHINIDIS N V. Derivative-free optimization: a review of algorithms and comparison of software implementations[J/OL]. *Journal of Global Optimization*, 2013, 56(3): 1247-1293. DOI:10.1007/s10898-012-9951-y.
- [2] ZHAO F, CHEN X, ZHU L. A hybrid numerical-symbolic solving strategy for equation-oriented process simulation and optimization[J/OL]. *AIChE Journal*, 2017, 63(7): 2764-2780. DOI:10.1002/aic.15622.
- [3] SHARMA N, LIU Y A. A hybrid science-guided machine learning approach for modeling chemical processes: A review[J/OL]. *AIChE Journal*, 2022, 68(5): e17609. DOI:10.1002/aic.17609.
- [4] YEE T F, GROSSMANN I E. Simultaneous optimization models for heat integration—II. Heat exchanger network synthesis[J/OL]. *Computers & Chemical Engineering*, 1990, 14(10): 1165-1184. DOI:10.1016/0098-1354(90)85010-8.
- [5] FISCHETTI M, JO J. Deep neural networks and mixed integer linear optimization[J/OL]. *Constraints*, 2018, 23(3): 296-309. DOI:10.1007/s10601-018-9285-6.
- [6] TSAY C, KRONQVIST J, THEBELT A, et al. Partition-Based Formulations for Mixed-Integer Optimization of Trained ReLU Neural Networks[C]//*Advances in Neural Information Processing Systems*: vol. 34. Curran Associates, Inc., 2021: 3068-3080[2024-05-22].
- [7] SCHWEIDTMANN A M, HUSTER W R, LÜTHJE J T, et al. Deterministic global process optimization: Accurate (single-species) properties via artificial neural networks[J/OL]. *Computers & Chemical Engineering*, 2019, 121: 67-74. DOI:10.1016/j.compchemeng.2018.10.007.
- [8] MARAGNO D, WIBERG H, BERTSIMAS D, et al. Mixed-integer optimization with constraint learning[J/OL]. *Operations Research*, 2023. DOI:10.1287/opre.2021.0707.
- [9] BIGGS M, HARISS R, PERAKIS G. Optimizing Objective Functions Determined from Random Forests[M]. Rochester, NY(2017-06-16)[2024-05-22]. DOI:10.2139/ssrn.2986630.
- [10] MIŠIĆ V V. Optimization of Tree Ensembles[J/OL]. *Operations Research*, 2020, 68(5): 1605-1624. DOI:10.1287/opre.2019.1928.
- [11] BERTSIMAS D, ÖZTÜRK B. Global optimization via optimal decision trees[J]. *Journal of Global Optimization*, 2023: 1-41.
- [12] BERTSIMAS D, MARGARITIS G. Global Optimization: A Machine Learning Approach[M/OL]. arXiv, 2023[2024-01-12]. <http://arxiv.org/abs/2311.01742>.
- [13] AMMARI B L, JOHNSON E S, STINCHFIELD G, et al. Linear model decision trees as surrogates in optimization of engineering applications[J/OL]. *Computers & Chemical Engineering*, 2023, 178: 108347. DOI:10.1016/j.compchemeng.2023.108347.
- [14] HORNİK K, STINCHCOMBE M, WHITE H. Multilayer feedforward networks are universal approximators[J/OL]. *Neural networks*, 1989, 2(5): 359-366. DOI:10.1016/0893-6080(89)90020-8.
- [15] GRIMSTAD B, ANDERSSON H. ReLU networks as surrogate models in mixed-integer linear programs[J/OL]. *Computers & Chemical Engineering*, 2019, 131: 106580. DOI:10.1016/j.compchemeng.2019.106580.
- [16] ANDERSON R, HUCHETTE J, MA W, et al. Strong mixed-integer programming formulations for trained neural networks[J/OL]. *Mathematical Programming*, 2020, 183(1): 3-39. DOI:10.1007/s10107-020-01474-5.
- [17] VIELMA J P, AHMED S, NEMHAUSER G. Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions[J/OL]. *Operations research*, 2010, 58(2): 303-315. DOI:10.1287/opre.1090.0721.
- [18] SCHWEIDTMANN A M, MITSOS A. Deterministic global optimization with artificial neural networks embedded[J/OL]. *Journal of Optimization Theory and Applications*, 2019, 180(3): 925-948. DOI:10.1007/s10957-018-1396-0.
- [19] YEE T F, GROSSMANN I E. Simultaneous optimization models for heat integration—II. Heat exchanger network synthesis[J/OL]. *Computers & Chemical Engineering*, 1990, 14(10): 1165-1184. DOI:10.1016/0098-1354(90)85010-8.
- [20] CHEN J J J. Comments on improvements on a replacement for the logarithmic mean[J/OL]. *Chemical Engineering Science*, 1987, 42(10): 2488-2489. DOI:10.1016/0009-2509(87)80128-8.