

基于课程正则化的物理信息神经网络渐进式训练策略

范黎林¹, 刘士豪¹, 李源^{1,2*}, 毛文涛^{1,2}, 陈宗涛¹

(1.河南师范大学计算机与信息工程学院, 河南 新乡 453007; 2.智慧商务与物联网技术河南省工程实验室, 河南 新乡 453007)

摘要:为降低物理信息神经网络(physics-informed neural networks, PINN)优化目标函数的复杂性和训练难度,提出一种基于课程正则化渐进式训练策略,在该策略中基于课程学习思想动态调整损失函数,使正则化项中偏微分方程所表征的物理信息从较平稳状态逐步过渡到变化剧烈状态,降低任务学习难度;加强损失函数中初始条件和边界条件部分的数据约束,平衡数据部分和物理信息部分损失;采用固定步长指数衰减学习率进行优化,尽可能避免目标函数陷入局部最小值。通过波动和热传导两类偏微分方程进行试验对比和分析,结果表明计算效率能够提升约50%,预测精度能够提高0.5~1个数量级。所提出方法可以有效提高PINN的数值稳定性和预测精度,加快PINN在复杂物理场学习任务中收敛速率。

关键词:物理信息神经网络;课程学习;损失函数;偏微分方程;数值稳定性

中图分类号:TP181

文献标志码:A

引用格式:范黎林,刘士豪,李源,等.基于课程正则化的物理信息神经网络渐进式训练策略[J].山东大学学报(工学版),2024,54(1):11-24.

FAN Lilin, LIU Shihao, LI Yuan, et al. Progressive training strategy of physics-informed neural networks based on curriculum regularization[J]. Journal of Shandong University (Engineering Science), 2024, 54(1):11-24.

Progressive training strategy of physics-informed neural networks based on curriculum regularization

FAN Lilin¹, LIU Shihao¹, LI Yuan^{1,2*}, MAO Wentao^{1,2}, CHEN Zongtao¹

(1. College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, Henan, China; 2. Engineering Lab of Intelligence Business & Internet of Things, Xinxiang 453007, Henan, China)

Abstract: A progressive training strategy based on curriculum regularization was proposed to reduce the complexity and training difficulty of the optimization objective function for physics-informed neural networks (PINN). In this strategy, the loss function was dynamically adjusted based on the idea of course learning. The physical information represented by the partial differential equation in the regularization term was gradually transitioned from a relatively stable state to a drastic state, which reduced the learning difficulty of task. The data constraints of initial conditions and boundary conditions in the loss function were strengthened to balance the losses between the data and physical information parts. A fixed step-size exponential decay learning rate was employed for optimization to avoid the objective function falling into the local minimum. Through experimental comparison and analysis of two types of partial differential equations, namely wave and heat conduction, the results showed that the computational efficiency could be improved by about 50%, and the prediction accuracy can be improved by an order of magnitude. The proposed method could improve the numerical stability and prediction accuracy of PINN, and accelerate the convergence rate of PINN in complex physical field learning tasks.

Keywords: physics-informed neural network; curriculum regularization; loss function; partial differential equation; numerical stability

0 引言

目前对于工程问题求解多采用物理模型驱动

的偏微分方程(partial differential equation, PDE)数值计算方法,包括有限差分法(finite difference method, FDM)^[1]、有限元法(finite element method, FEM)^[2]、有限体积法(finite volume method,

收稿日期:2023-07-03

基金项目:国家自然科学基金资助项目(11702087,U1704158);河南省科技攻关重点项目(212102210103)

第一作者简介:范黎林(1970—),男,河南周口人,副教授,硕士生导师,博士,主要研究方向为智能商务、智能交通。E-mail:fanlilin@htu.edu.cn

*通信作者简介:李源(1989—),女,河南新乡人,副教授,硕士生导师,博士,主要研究方向为计算力学、机器学习、边界元法。

E-mail:liyuan2015097@163.com

FVM)^[3]、边界元法(boundary element method, BEM)^[4]等。近年来,随着人工智能理论在图像识别^[5]、自然语言处理^[6]、医学图像处理^[7]、科学计算^[8]等领域快速发展,深度学习开始引入到偏微分方程数值求解领域。文献[9]提出的物理信息神经网络(physics-informed neural networks, PINN),将偏微分方程所表征的物理信息融入到损失函数构造中,采用数据驱动和物理驱动相结合方式训练出能够自动满足偏微分方程解的神经网络模型。受益于深度学习强大表达能力,PINN可通过有限数据点自动学习到隐藏的物理信息,克服传统数值计算方法对物理先验知识和网格的依赖性,提供一种更加灵活的解决方案。

PINN在求解PDE时模型前处理成本较低、统一了PDE正问题和反问题的计算框架,能够在数值模型动态运行过程中自然融入新的观测数据,有利于数据同化实施。目前广泛应用在多种物理问题PDE求解中,如热传导方程^[10]、Navier-Stokes方程^[11]、Poisson方程^[12]、对流方程^[13]等。PINN方法存在着计算效率低和计算性能不稳定问题,学者们提出了不同的改进策略。文献[14-15]在神经网络中通过修改每个隐藏层中激活函数斜率,提高了训练速度,随后又通过在激活函数中引入缩放因子,与神经网络各参数共同训练更新,进一步加快了网络的训练速度;文献[16]通过传统的分离变量法获取了热传导方程的物理解在时间和空间上周期性、指数衰减性函数,并融入到了激活函数的设计当中,提高了收敛速度;文献[17]提出了学习率退火算法,在网络模型训练过程中自适应地为各项损失函数分配最佳权重,使得训练精度提高;文献[18]通过构建距离函数,使PINN的输出自动满足初始及边界条件,损失函数中只包含物理信息损失部分,是一种称为“硬约束”的PINN,提高了PINN求解精度;文献[19]提出了梯度优化型PINN,通过梯度统计来平衡模型训练过程中不同损失函数项之间相互作用,从而使PINN在梯度出现反复振荡、预测精度不稳定时具有更强的数值稳定性;PINN变种例,如变分型PINN^[20]和守恒型PINN^[21]等通过将能量守恒原理嵌入到损失函数中,能够有效加速迭代过程收敛,计算精度有所降低。随着PINN快速发展,多个用于科学计算的机器学习库相继提出,如文献[22]开发的DeepXDE能够实现求解不同的偏微分方程,可以根据需求自定义物理求解域;类似还有ADCME^[23]、Sciann^[24]等科学计算库发布用于求解不同类型PDE正反问题。

以上方法主要针对PINN收敛速度或求解精度进行改进,但PINN数值稳定性也是制约其发展重要瓶颈。即使对于同一类PDE,PINN求解精度仍然会受到PDE中系数影响,具体表现在:当方程中系数相对较小时,PINN能够得到较高计算精度;当方程中的系数相对较大时,PINN计算误差随之增大,甚至出现错误结果。文献[25]分析了同一类PDE在不同系数下PINN求解精度和损失景观,发现PINN求解误差较大并不是由于神经网络的表达能力导致的,而是由于系数增大时损失景观变得更加复杂,使得目标函数优化更加困难所导致的,可以通过逐步改变PDE损失项所表征的物理信息,调整PINN学习难度解决此类问题。本研究借助课程学习的思想,提出一种课程正则化渐进式训练策略来降低PINN学习难度。

1 相关知识

1.1 深度神经网络与自动微分

近年来,深度神经网络(deep neural networks, DNN)^[26]以其卓越表达能力,在自然语言处理、图像识别等领域得到广泛应用。前馈全连接神经网络可以假设为输入层、多个隐藏层和输出层的堆叠。两个相邻层之间的连接,例如从第*i*-1层到第*i*层,可以表示如下:

$$\mathbf{z}_i = \sigma(\mathbf{b}_i + \mathbf{w}_i \mathbf{z}_{i-1}), \quad 1 \leq i \leq n, \quad (1)$$

$$\mathbf{z}_{n+1} = \mathbf{b}_{n+1} + \mathbf{w}_{n+1} \mathbf{z}_n, \quad (2)$$

式中,*n*为隐藏层的总数, $\sigma(\cdot)$ 表示激活函数, \mathbf{z}_n 和 \mathbf{z}_{n+1} 分别为输入和输出张量, \mathbf{w}_i 和 \mathbf{b}_i 为第*i*层中的可训练权重和偏置。

使用DNN作为偏微分方程解的预测模型。为了将物理信息嵌入神经网络训练中,需要构造与PDE相关的损失函数。PDE中存在物理量关于时空变量的偏导数,需要借助神经网络自动微分^[27]才能实现损失函数计算。如图1所示,假设网络将变量*x*作为输入,经过非线性变换之后输出*f*。 $y_j^{(i)}$ 表示第*i*个隐层第*j*个神经元的输出。由于输出*f*可以表示为*x*的嵌套函数,则可以应用链式法则来计算*f*关于*x*的导数,如:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y_1^{(2)}} \frac{\partial y_1^{(2)}}{\partial y_1^{(1)}} \frac{\partial y_1^{(1)}}{\partial x} + \frac{\partial f}{\partial y_2^{(2)}} \frac{\partial y_2^{(2)}}{\partial y_2^{(1)}} \frac{\partial y_2^{(1)}}{\partial x} + \frac{\partial f}{\partial y_2^{(2)}} \frac{\partial y_2^{(2)}}{\partial y_1^{(1)}} \frac{\partial y_1^{(1)}}{\partial x} + \frac{\partial f}{\partial y_1^{(2)}} \frac{\partial y_1^{(2)}}{\partial y_1^{(1)}} \frac{\partial y_1^{(1)}}{\partial x}, \quad (3)$$

式中等号右端第一项对应于图1中的虚线路径。

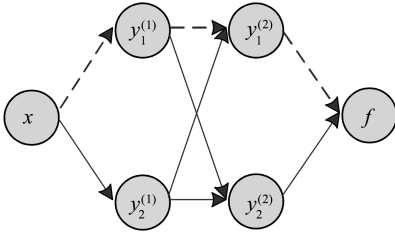


图 1 自动微分示意图

Fig.1 Automatic differentiation diagram

1.2 PINN 方法概述

PINN 与常规深度神经网络不同之处在于, PINN 能够将物理信息嵌入到网络中,使网络输出更加符合物理规律。偏微分方程是描述物理规律常用的形式,为描述 PINN 嵌入物理信息的方法,考虑如下形式的偏微分方程:

$$\begin{cases} u_t + N[u; \lambda] = 0, & x \in \Omega, t \in [0, T] \\ I(u; \lambda) |_{(x, t_0)} = 0, & x \in \Omega \\ B(u; \lambda) |_{(x, t)} = 0, & x \in \partial\Omega, t \in [0, T] \end{cases}, \quad (4)$$

式中: u 表示式(4)待求解的物理量,是关于 (x, t) 的函数; x, t 分别是空间变量和时间变量; u_t 表示 u 对 t 的偏导数; $N[u; \lambda]$ 表示对 u 进行微分运算的非线性算子, λ 为方程中的参数; Ω 表示欧氏空间 \mathbf{R}^D 的子集; $\partial\Omega$ 表示 Ω 的边界; T 表示时间终止时刻; $I(u; \lambda)$ 和 $B(u; \lambda)$ 分别表示方程的初始条件(initial condition, IC)和边界条件(boundary condition, BC)。在已知方程参数 λ 的情况下,输入时空坐标 (x, t) , 可以求解出任意时空点的 $u(x, t)$ 。

PINN 算法利用神经网络作为函数逼近器来近似式(4)的解 $u(x, t)$ 。PINN 的基本框架如图 2 所示,包含了神经网络逼近函数(如左侧虚线框所示)和信息约束(如右侧虚线框所示),其中信息约束又分为物理信息约束(即 PDE 约束)和数据约束(即 BC&IC 约束)两部分。

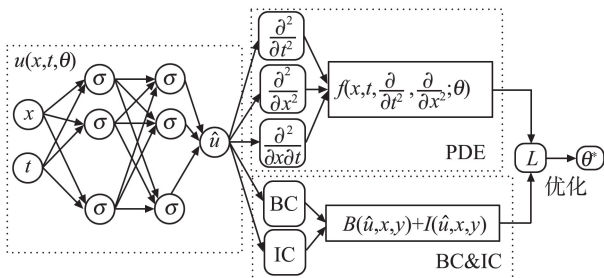


图 2 PINN 框架

Fig.2 PINN framework

PINN 求解过程如下:

(1) 在时空域内采样生成 3 组数据点作为训练数据,分别为边界数据点、初始数据点和域内数

据点;

(2) 构建神经网络模型 $\hat{u}((x, t), \theta)$, θ 为初始网络参数;

(3) 根据偏微分方程、边界条件和初始条件,采用自动微分技术获取物理信息残差和数据残差,将其作为各损失项构建损失函数;

(4) 将训练数据带入网络模型进行训练,以最小化损失函数为目标来更新神经网络的参数 θ ,使训练后的神经网络能够模拟出偏微分方程的解 $u(x, t)$ 。PINN 的损失函数为:

$$L = L_f + L_u, \quad (5)$$

式中: $L_f = \frac{1}{N_f} \sum_{i=1}^{N_f} |f(t_f^i, x_f^i)|^2$, L_f 为物理信息部分的损失,其中 $f(t_f^i, x_f^i)$ 表示 PDE 的残差, $\{t_f^i, x_f^i\}_{i=1}^{N_u}$ 表示 PDE 的域内数据点; L_u 为数据部分的损失, $L_u = \frac{1}{N_I} \sum_{i=1}^{N_I} |u(t_i^I, x_i^I) - \hat{u}_i^I|^2 + \frac{1}{N_B} \sum_{i=1}^{N_B} |u(t_B^i, x_B^i) - \hat{u}_B^i|^2$, 包括初始值损失和边界值损失,其中 $u(t_i^I, x_i^I)$ 和 \hat{u}_i^I 分别表示初始时刻真实值和预测值, $u(t_B^i, x_B^i)$ 和 \hat{u}_B^i 分别表示边界部分真实值和预测值。PINN 的损失函数中,不仅仅依靠数据约束,还施加了物理信息约束。

2 基于课程正则化的渐进式训练策略

由于 PINN 损失函数需要表征各类偏微分方程知识,优化目标复杂, PINN 训练过程难以收敛、预测性能不稳定。即使对于同一类 PDE,方程中系数改变时 PINN 计算精度也会受到显著影响。文献[25]分析了同一类 PDE 在不同系数下 PINN 求解精度和损失景观,发现方程系数较大时 PINN 求解误差较大,这是由于系数增大时损失景观变得更加复杂,使得目标函数优化更加困难所导致的。

课程学习作为一种训练策略,由 Montreal 大学的 Bengio 教授团队^[28]提出,主要思想是模仿人类学习过程,模型从简单任务开始学习,逐步进阶到复杂任务,可以引导训练更容易走向全局最优解,使模型具有更好泛化性能。从模型优化角度看,课程学习提供了如图 3 所示的一种优化目标序列,先从底部较为平滑的目标开始优化,这样能够快速找到全局最小值,在整个训练过程中跟踪最小值。除此之外,从更容易的目标中学习到的最小值具有更好泛化能力,更有机会接近目标函数

的全局最小值,降低了目标函数的优化难度。

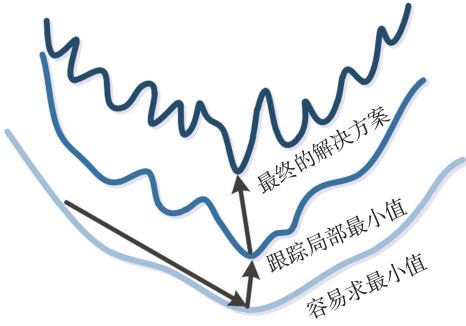


图3 课程学习示意图
Fig.3 Diagram of curriculum learning

2.1 基于课程正则化的损失函数动态调整策略

基于课程正则化的损失函数动态调整策略旨在通过调整训练过程中损失函数降低优化难度。在该策略中,预先根据数据驱动部分损失进行训练,待该部分损失达到预设容差后再加入物理信息部分正则化项,同时正则化项中 PDE 系数按照设定策略从较小值逐步增加至目标值,降低学习任务的难度;通过调整不同损失项之间权重配比,加强初边值条件约束,实现各部分损失的平衡。该策略的框架如图4所示。

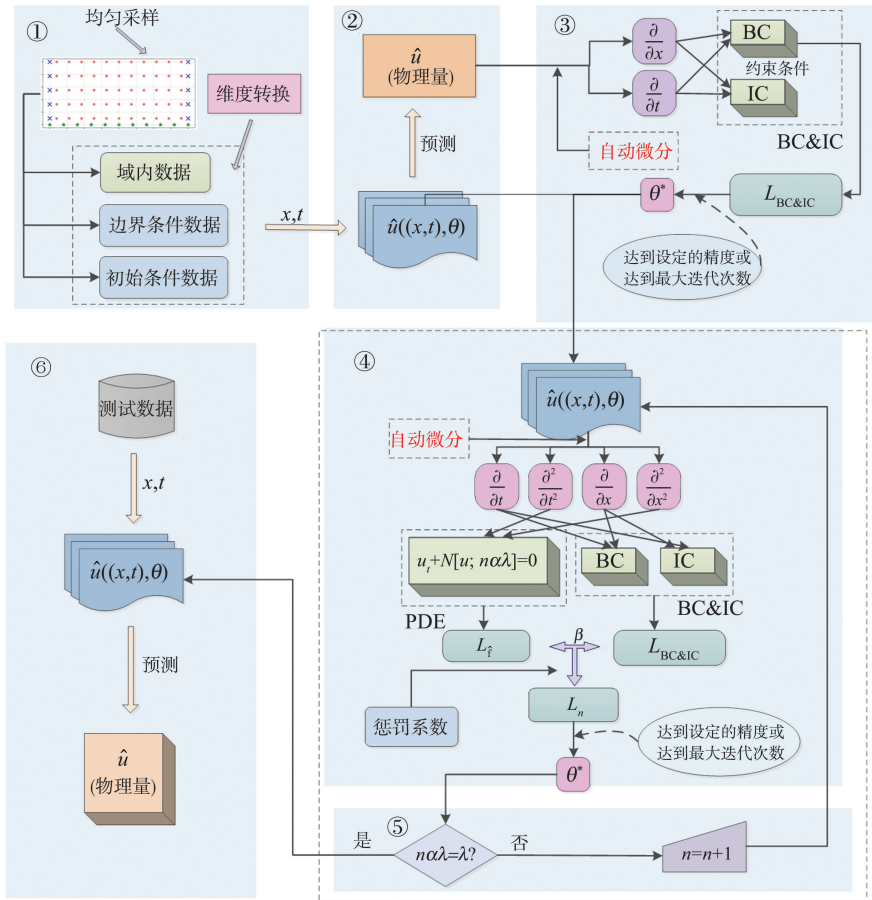


图4 基于课程正则化的渐进式训练策略
Fig.4 A progressive training strategy with curriculum regularization

损失函数动态调整策略的主要步骤如下:

(1) 采集训练数据。训练数据采集方式有均匀采样、拉丁超立方采样等。以均匀采样为例,在规定时空域中进行均匀采样,将获取的 (x, t) 二维矩阵集进行笛卡尔乘积,得到均匀分布的采样点矩阵,从中划分初始条件部分、边界条件部分以及域内采样点数据集,并对各数据集数据进行缩放处理;

(2) 通过 Pytorch 深度学习框架构建全连接神经网络模型 $\hat{u}((x, t), \theta)$, θ 为初始网络参数,得到

预测物理量 \hat{u} ;

(3) 根据所求偏微分方程边界条件和初始条件构造损失函数 $L_{BC\&IC}$,将训练数据带入网络模型进行训练,以最小化损失函数为目标来更新神经网络参数 θ ,可以采用 SGD、Adam^[29]、L-BFGS^[30] 等优化器进行优化,本研究采用 Adam 优化器;

(4) 待 $L_{BC\&IC}$ 达到预设容差或最大迭代次数后加入物理信息部分残差,构造新的损失函数 $L_{BC\&IC} + L_f$, L_f 为物理信息部分损失;引入惩罚系数 β 调整不同损失项之间权重配比,旨在加强初边值部分数据

约束,平衡各损失项对结果的影响,使结果尽可能满足设定的初边值条件;采用课程正则化策略使 L_1 损失项中系数 λ 从较小值开始逐步增加至目标值。本研究采用 n 和 α 作为控制参数来调整系数 λ 的值,其中 $n=1,2,\dots,N,\alpha=1/N,N$ 代表设置的课程正则化步数;

(5) 在当前系数下训练结束后判断 $n\alpha=1$ 是否为真,如果不为真则使 $n=n+1$ 继续训练,否则停止训练;

(6) 训练得到 PINN 的模型,将测试数据带入到模型中,输出所求偏微分方程预测解 $\hat{u}((x,t),\theta^*)$ 。

第(3)~(5)步中损失函数动态调整的具体流程如算法1所示。

算法1: 基于课程正则化的渐进式训练策略

Input: 边界数据 A_1 , 初始数据 A_2 , 域内数据 A_3

Output: \hat{u}

while $n\alpha \leq 1$ do

 if $n=1$ then

 for $\gamma \leftarrow 0$ to γ_{\max} do

 if $\gamma \leq \gamma_{\max}$ && $L_1 > v$ then

 以 L_1 为目标函数更新参数

 elif $L_2 > c$ then

 以 L_2 为目标函数更新参数

 else

 terminate

 end for

 else

 for $\gamma \leftarrow 0$ to γ_{\max} do

 if $L_2 > h$ then

 以 L_2 为目标函数更新参数

 else

 terminate

 end for

 end if

$n=n+1$

$\hat{\lambda}=n\alpha\lambda$

end while

算法1中, A_1 为边界条件部分的数据集, A_2 为初始条件部分数据集, A_3 为域内采样点数据集, \hat{u} 为输出预测解, γ 为迭代次数, v,c,h 分别为预设损失值大小,初边值部分构成损失函数 L_1 为:

$$L_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} |u(t_1^i, x_1^i) - \hat{u}_1|^2 +$$

$$\frac{1}{N_B} \sum_{i=1}^{N_B} |u(t_B^i, x_B^i) - \hat{u}_B|^2, \quad (6)$$

初边值部分和物理信息部分共同构成损失函数 L_2 为:

$$L_2 = \beta \left(\frac{1}{N_1} \sum_{i=1}^{N_1} |u(t_1^i, x_1^i) - \hat{u}_1|^2 + \frac{1}{N_B} \sum_{i=1}^{N_B} |u(t_B^i, x_B^i) - \hat{u}_B|^2 \right) + \frac{1}{N_f} \sum_{i=1}^{N_f} |f(t_f^i, x_f^i, \hat{\lambda})|^2, \quad (7)$$

式中, β 为损失函数的惩罚系数,用于调整损失项之间权重配比。

2.2 基于指数衰减学习率的 PINN 优化方案

学习率作为监督学习以及深度学习中重要超参数,决定着目标函数能否收敛到局部最小值以及何时收敛到最小值。合适的学习率能够使目标函数在合适时间内收敛到全局最小值。与纯数据驱动的深度学习方法不同,PINN 中还包含了物理信息约束,使得损失函数景观更加复杂,目标函数容易陷入局部最小值而过早收敛。Adam 优化器采用了学习率自适应机制,学习率可变化范围较小,当初始学习率较小时容易在训练前期陷入局部最小值,当初始学习率较大时容易在训练后期引起震荡而无法收敛。本研究在 Adam 优化器的基础上引入学习率衰减机制,采用固定步长指数衰减学习率更新参数,每经过 K 次迭代,模型参数 θ 更新为:

$$\theta_{\text{new}} = \theta_{\text{old}} - \frac{\eta r^n}{\sqrt{v_t + \varepsilon}} m_t, \quad (8)$$

式中: η 代表初始学习率; $r \in (0,1)$ 代表衰减率; v_t 代表二阶矩估计; m_t 代表一阶矩估计; ε 代表数值稳定项; $n=0,1,2,\dots,N$ 为当前衰减次数,即每经过 K 次迭代 n 的值加1,学习率衰减为原来的 r 倍。

图5(a)展示了固定步长下指数衰减学习率随迭代次数变化曲线,取 $K=500,r=0.98$ 。从图5(a)中可以看出:衰减机制的引入使得学习率随着迭代次数增加逐步减小。训练前期采用较大学习率可以帮助模型跳出局部最优解,防止模型过拟合,提高泛化能力;训练后期采用逐步减小的学习率,可以使训练更加关注学习任务的细节,提高训练精度。图5(b)对比了两种学习率下目标函数损失下降曲线,可以看出:在相同迭代次数下,默认学习率损失下降虽然震荡幅度较小,但是损失更易陷入局部最小值而过早收敛,无法达到更高的训练精度。固定步长指数衰减学习率则能使目标函数跳出局部最优解并达到更高的训练精度。

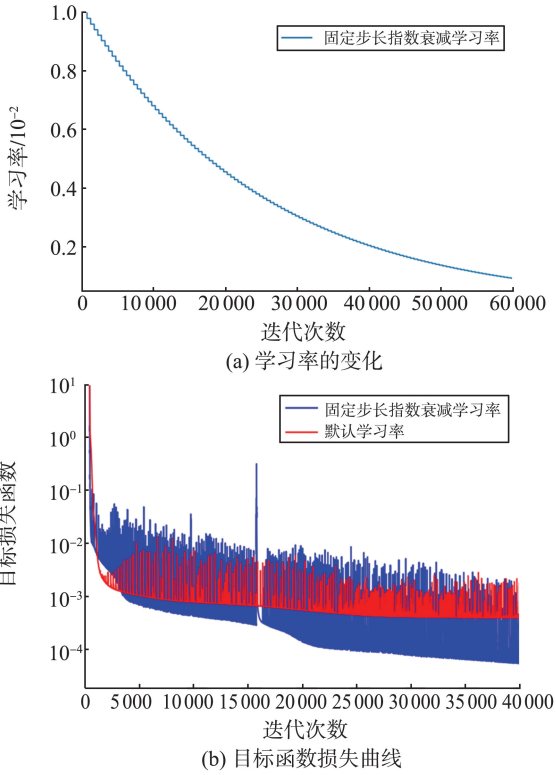


图5 学习率的变化和不同学习率下的损失下降曲线
Fig.5 Variations of learning rate and loss reduction under different learning rates

3 试验分析

通过波动方程和热传导方程数值计算试验说明训练数据的采集方法,对比常规 PINN 和本研究所提出的课程正则化 PINN (curriculum regularization PINN, CRPINN) 的预测精度和计算效率,验证各项改进措施的有效性。

3.1 数据采集

配置点的整个集合可以由 $S = \{S_D, S_{BC}, S_{IC}\}$ 表示, S 由整个时空域的域内时空坐标 S_D 、边界条件 S_{BC} 和初始条件 S_{IC} 组成,如下所示:

$$\begin{cases} S_D = \{(x, t) | x \in \Omega, t \in [0, T]\} \\ S_{BC} = \{(x, t) | x \in \partial\Omega, t \in [0, T]\}, \\ S_{IC} = \{(x, t=0) | x \in \Omega\} \end{cases} \quad (9)$$

式中, Ω 表示欧氏空间 \mathbf{R}^D 的子集, $\partial\Omega$ 表示时空域的边界, x, t 分别是空间变量和时间变量, T 是时间终止时刻。训练数据通过均匀采样策略生成,如图 6 所示。

采用 L_2 相对误差评估 PINN 的预测性能。在 L_2 相对误差中,通过计算两个向量之间的 L_2 范数差异,来衡量这两个向量之间的相对误差为:

$$L_2 = \frac{1}{N} \sum_{i=0}^N \frac{\|\hat{u} - u\|^2}{\|u\|^2}, \quad (10)$$

式中, N 是采样点总数, u 为精确解, \hat{u} 为 PINN 预测解。

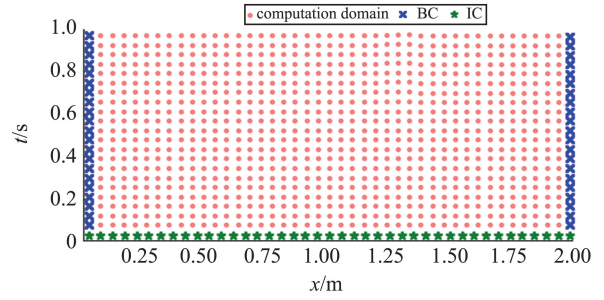


图6 试验数据的采样示意图

Fig.6 Schematic diagram of experimental data sampling

3.2 波动方程试验

在波动方程问题上,假设空间域 x 取值 $[0, 2]$, 时间 t 取值 $[0, 1]$ 。通过均匀采样在时空域上生成 20 000 个 (x, t) 数据点,其中在时间 t 上均匀离散 100 个点,在空间 x 上均匀离散 200 个点。从中取出 200 个点作为初始条件部分训练数据、400 个点作为边界条件部分训练数据,其余为域内配置点。使用 3 层全连接神经网络,每层 64 个神经元,采用双曲正切激活函数 (tanh),使用 Adam 优化器和固定不长指数衰减学习率来训练网络。

考虑一维波动方程

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \quad x \in [0, 2], t \in [0, 1], \quad (11)$$

满足边界条件:

$$u(0, t) = 0, \quad (12)$$

$$\frac{\partial u}{\partial x}(2, t) = 0, \quad (13)$$

和初始条件:

$$u(x, 0) = 0.005x, \quad (14)$$

$$\frac{\partial u}{\partial t}(x, 0) = 2.5, \quad (15)$$

式中: $u(x, t)$ 是待求解的位移函数, c 表示波速,在均匀介质中 c 为常数。

方程(11)的精确解为:

$$u(x, t) = \sum_{n=1,3,5,\dots}^{\infty} \left\{ \frac{4}{n\pi} \left[(-1)^{\frac{n-1}{2}} \frac{0.01L}{n\pi} \right] \cos \frac{cn\pi}{2L} t + \frac{20L}{cn^2\pi^2} \sin \frac{cn\pi}{2L} t \right\} \sin \frac{n\pi x}{2L}, \quad (16)$$

式中 $L=2$ 。

3.2.1 预测精度对比分析

波动方程精确解、常规 PINN 预测解和 CRPINN 预测解的对比如图 7 所示。从图 7 可以看出,随着波速 c 的增加,位移解 u 的分布越来越复杂,呈现周期性振动特点。当波速 $c^2 = 1$ 时,常规 PINN 和 CRPINN 的预测解均与精确解吻合较好,

随着波速的增加,可以明显看出常规 PINN 预测解的分布与精确解相差越来越大。这是由于波速变大时位移解 u 越来越复杂,学习难度也随之增加。与常规 PINN 相比,所提出的 CRPINN 预测解能够与精确解更好地吻合,结果精度显著提升。

为了进一步说明优化前后 PINN 预测精度变化,绘制预测解与精确解之间绝对误差的绝对值 (Δu) 分布,如图 8 所示。从图 8 中可以看出,随着波速 c 的增加,常规 PINN 的预测误差明显高于 CRPINN 预测误差。

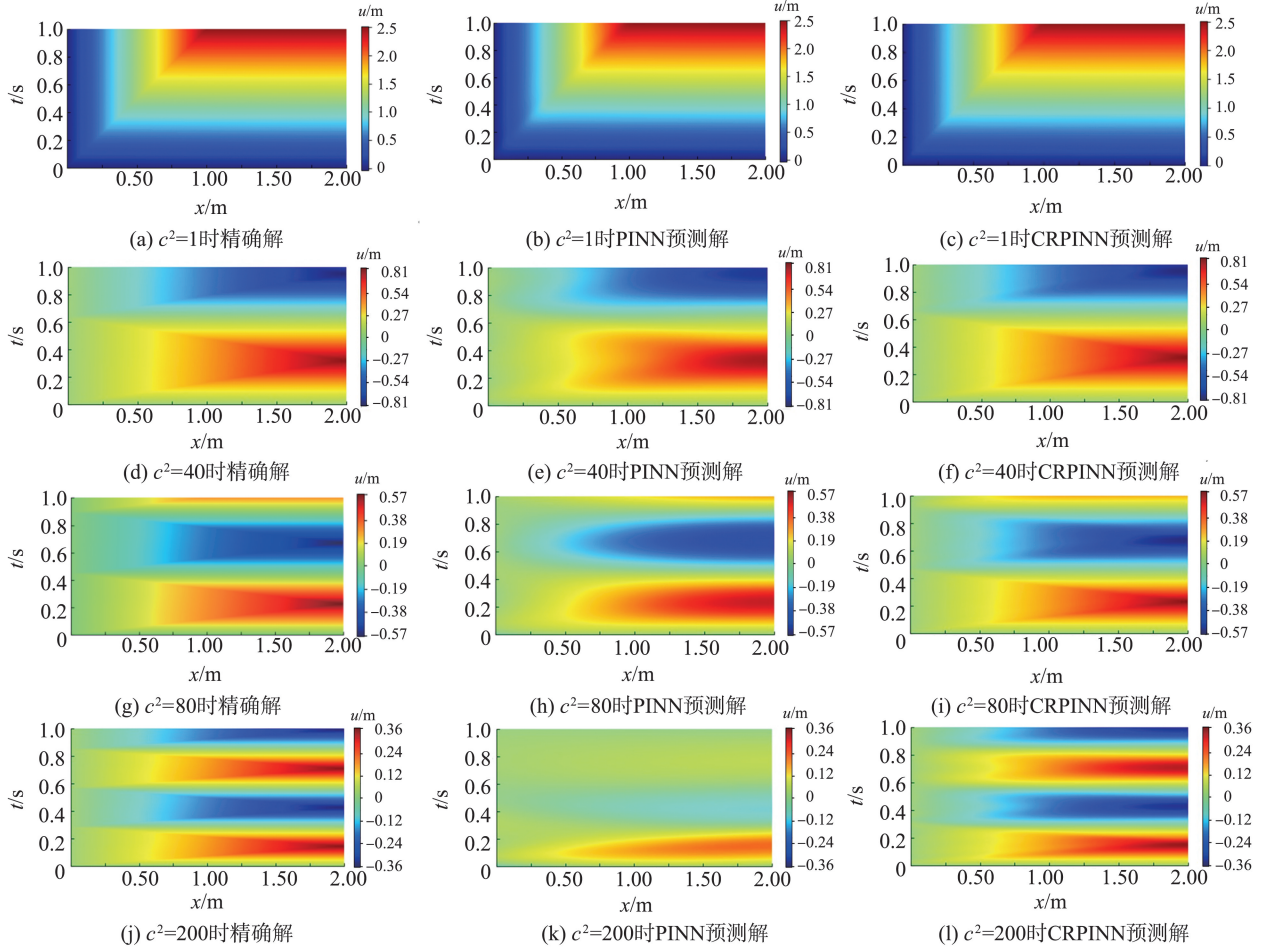
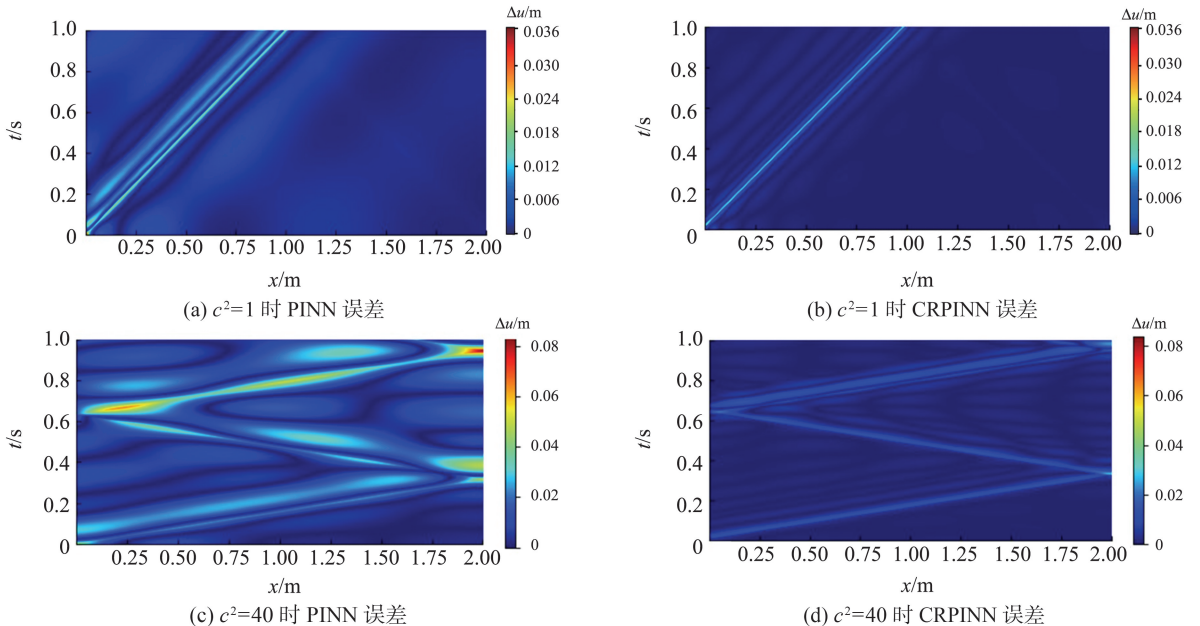


图 7 不同波速下波动方程的预测解与精确解对比图

Fig.7 Comparison chart of predicted solutions and exact solutions of the wave equation under different wave speeds



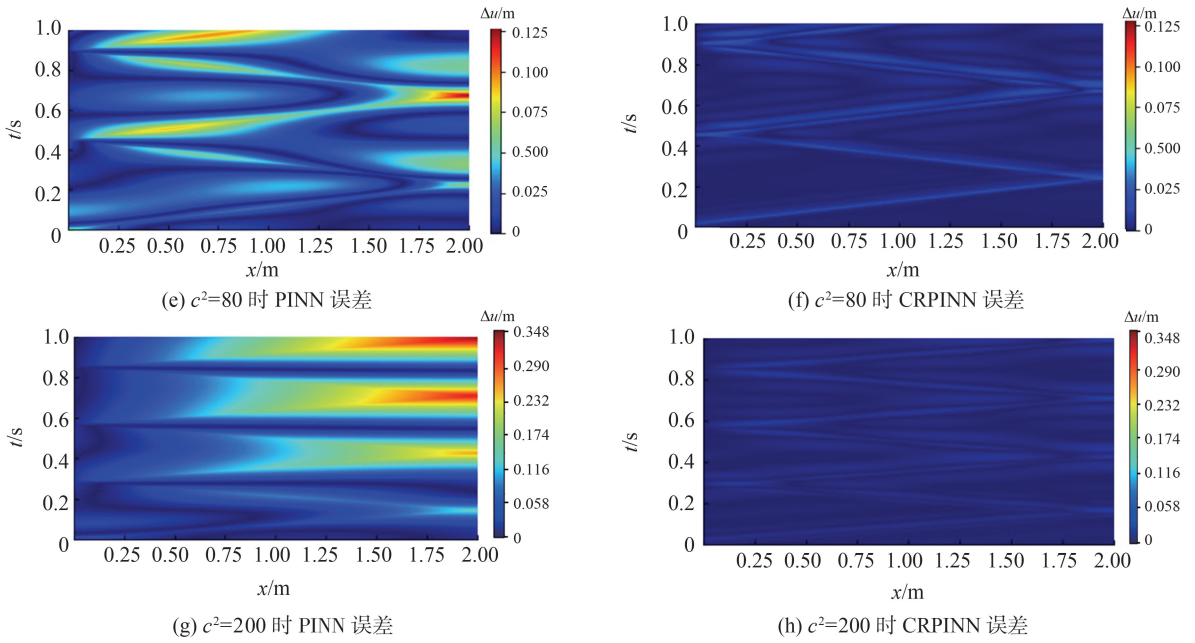


图8 不同波速下波动方程预测结果绝对误差的绝对值对比图

Fig.8 Comparison of absolute value of predicted absolute error in wave equation results under different wave velocities

图9展示了在 $x=1$ 位置上波动方程的解随时间的变化曲线,可以看出当 $c^2=1$ 时,常规 PINN 和 CRPINN 的结果曲线均与精确解吻合较好,但是随

着波速的增加,常规 PINN 结果曲线与精确解相差越来越大。与常规 PINN 相比,CRPINN 预测结果曲线与精确解的一致性更高。

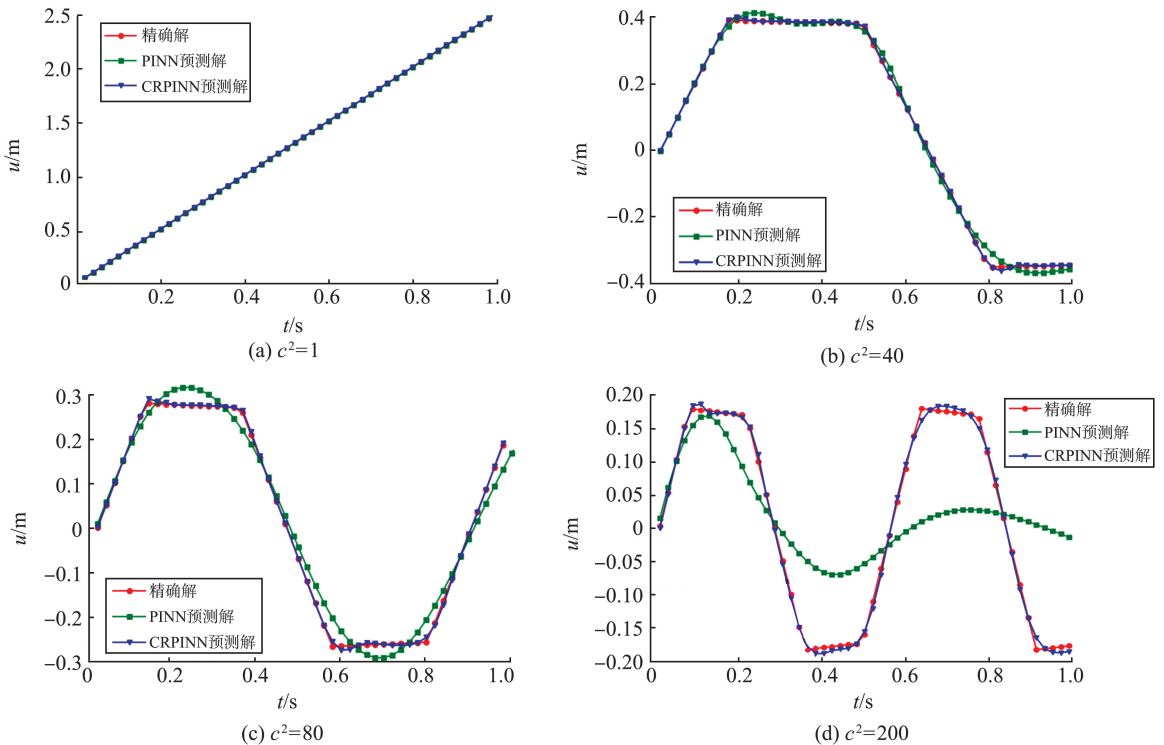


图9 当 $x=1$ 时波动方程的解随时间的变化曲线对比图

Fig.9 Comparison plot of the solution of the wave equation with time for $x=1$

试验中发现,PINN 算法中数据部分损失和物理信息部分损失之间权重占比对结果有较大影响。如式(7)所示,当引入惩罚系数 β 加强初边值部分数据约束时,可使结果尽可能严格满足设定

的初边值条件,提高预测精度。通过试验对比了几种取值情况,发现当惩罚系数 $\beta=100$ 时 PINN 的预测精度较高。为了研究惩罚系数、学习率以及损失函数动态调整 3 种措施对结果的影响,设

计 3 种不同的试验验证,每种措施作用分别为:

试验 1 在常规 PINN 的基础上引入惩罚系数 $\beta=100$,平衡数据部分损失和物理信息部分损失,加强初边值条件的数据约束,验证该措施对预测精度的影响;

试验 2 在试验 1 的基础上,采用固定步长指数衰减学习率来进行优化,验证学习率对预测精度的影响。本试验采用 Adam 优化器,初始学习率为

表 1 波动方程在不同波速、不同训练设定下的 L_2 相对误差

Table 1 L_2 relative error of the wave equation under different wave speeds and different training settings

波速平方 c^2	L_2 相对误差			
	常规 PINN	PINN+惩罚系数	PINN+惩罚系数+指数衰减学习率	PINN+惩罚系数+指数衰减学习率+课程正则化
1	0.002 108	0.001 367	0.001 100	0.000 995
40	0.044 476	0.021 149	0.016 027	0.010 869
80	0.136 928	0.053 161	0.025 263	0.019 526
200	0.728 209	0.151 789	0.076 639	0.035 905

从表 1 中可以看出:(1) 引入惩罚系数 $\beta=100$ 使得测试误差显著下降,说明惩罚系数能够有效平衡数据部分和物理信息部分损失;(2) 指数衰减学习率引入进一步降低了预测误差,说明该措施可以避免在优化前期陷入局部最小值,并最终以较高精度接近全局最小值;(3) 课程正则化策略的加入进一步降低了预测误差,说明在简单任务的基础上进

0.01,固定步长为 500 步,衰减率为 0.98;

试验 3 在试验 2 的基础上,加入课程正则化方法,验证该措施对预测精度影响。本试验中课程正则化步数 N 取值为 10。当 $c^2=1$ 时,课程正则化策略中系数 c^2 从 0.1 开始逐步增加至目标值;当 $c^2>1$ 时,课程正则化策略中系数 c^2 从 1 开始逐步增加至目标值。

表 1 给出了以上不同试验设置下 PINN 预测解的 L_2 相对误差。

行训练能够有效降低复杂任务的学习难度,并取得更高预测精度。

3.2.2 计算效率对比分析

为了衡量改进后 PINN 方法对计算效率影响,对比了常规 PINN 方法和所提出 CRPINN 方法训练损失迭代曲线,如图 10 所示。

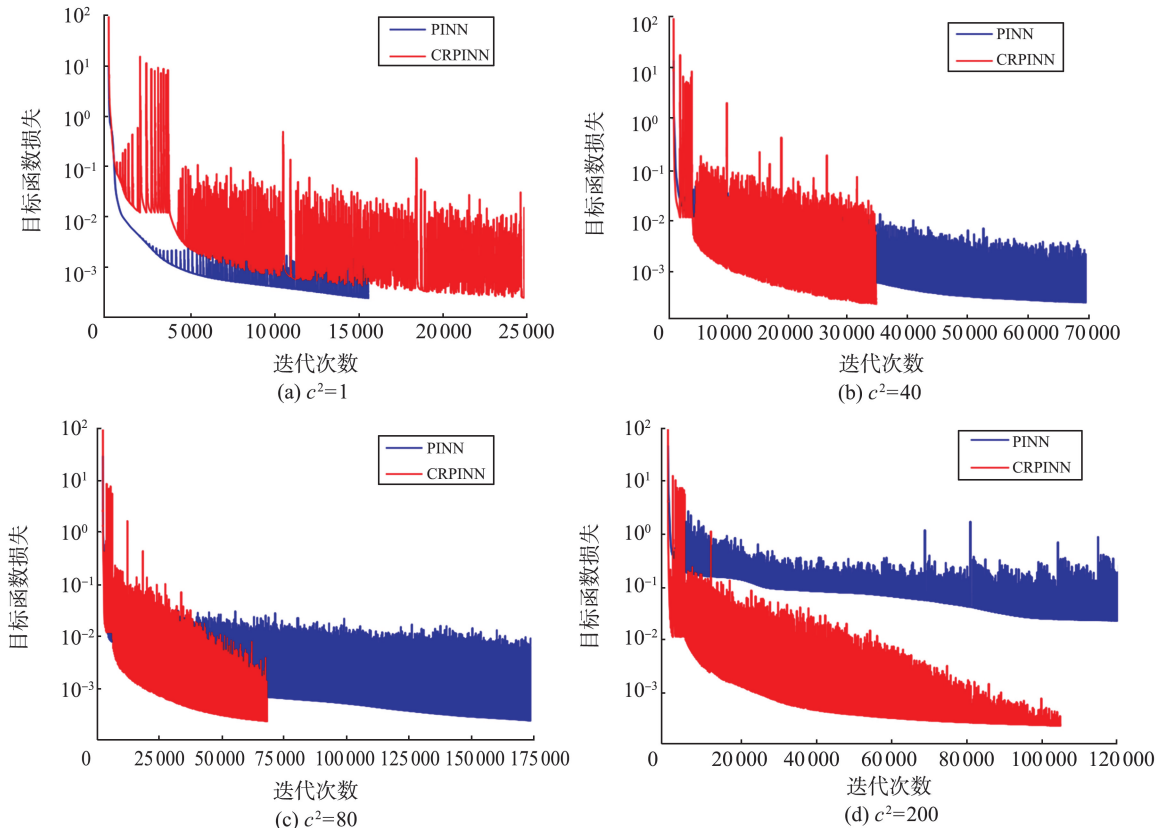


图 10 常规 PINN 方法与 CRPINN 方法的计算效率对比

Fig.10 Comparison of computational efficiency between conventional PINN method and CRPINN method

从图 10 可以看出:

(1) 当 $c^2 = 1$ 时, 常规 PINN 方法比 CRPINN 更快下降到了程序中设置预期误差, 说明常规 PINN 方法可以更快完成简单任务学习, 但从表 1 可以看出, 常规 PINN 方法在测试集上的 L_2 相对误差明显高于 CRPINN, 说明 CRPINN 具有更好泛化性能;

(2) 当 $c^2 = 40$ 和 $c^2 = 80$ 时, CRPINN 方法训练时收敛速度明显高于常规 PINN 方法, 能够更快下降到程序中设置的预期误差。从图 10(b)(c) 中的损失迭代曲线可知 CRPINN 的计算效率相较于常规 PINN 能够提升约 50%, 说明所提出渐进式训练策略起到了课程学习作用, 能够引导训练模型更快收敛到最优解;

(3) 当 $c^2 = 200$ 时, 常规 PINN 方法已无法收敛到程序中设置的预期误差, 在测试集上的 L_2 相对误差高达 72%。而 CRPINN 仍然可以收敛到预期误差, 说明所提出的改进措施有效提高了 PINN 数值稳定性。

3.3 热传导方程试验

试验设置同波动方程一致, 对于一维热传导方程

$$\frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2}, \quad x \in [0, 2], \quad t \in [0, 1], \quad (17)$$

满足边界条件:

$$u(0, t) = u(2, t) = 0, \quad (18)$$

和周期性(正弦)初始条件:

$$u(x, 0) = \sin(\pi x), \quad (19)$$

式中, $u(x, t)$ 是待求解的物理量, k 表示扩散系数。

该问题精确解为:

$$u(x, t) = e^{-\pi^2 k t} \sin(\pi x). \quad (20)$$

热传导方程精确解、常规 PINN 预测解和 CRPINN 预测解对比如图 11 所示。从图 11 中可以看出, 当扩散系数 $k = 0.1$ 时可以明显观察到计算域内温度随时间的变化, 而随着 k 增大温度逐渐趋于平稳, 这是因为在 $[0, 1]$ 时间内, 扩散系数越大, 温度扩散越快。常规 PINN 的预测解和 CRPINN 的预测解均能反应出温度变化的趋势。

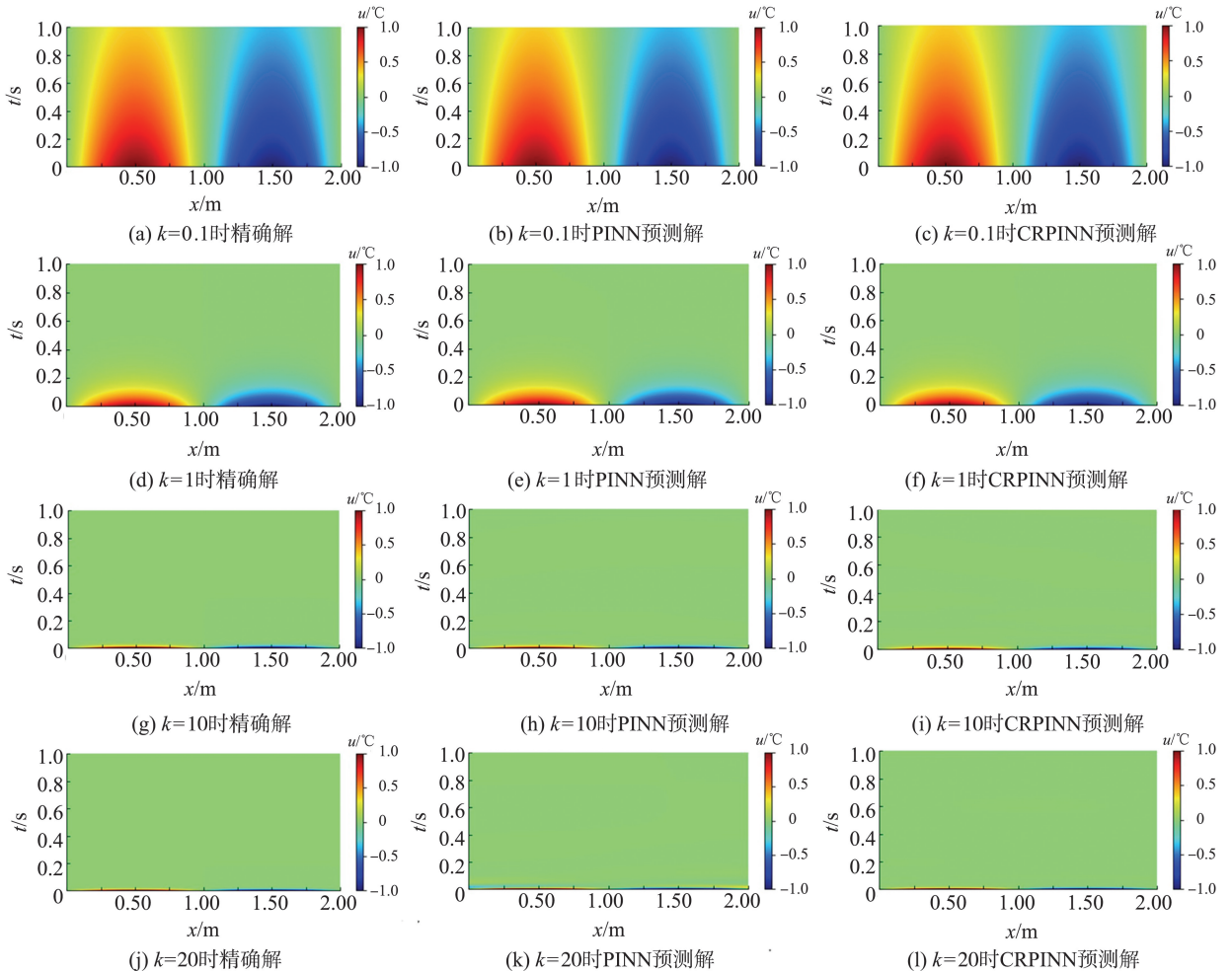


图 11 不同扩散系数下热传导方程的预测解与精确解对比图

Fig. 11 Comparison chart of predicted solutions and exact solutions of the heat conduction equation under different diffusion coefficients

为了进一步说明优化前后 PINN 预测精度变化,绘制预测解与精确解之间的绝对误差的绝对值分布如图 12 所示。从图 12 中可以看出,随着扩散

系数 k 增加,CRPINN 预测误差明显低于常规 PINN 预测误差。

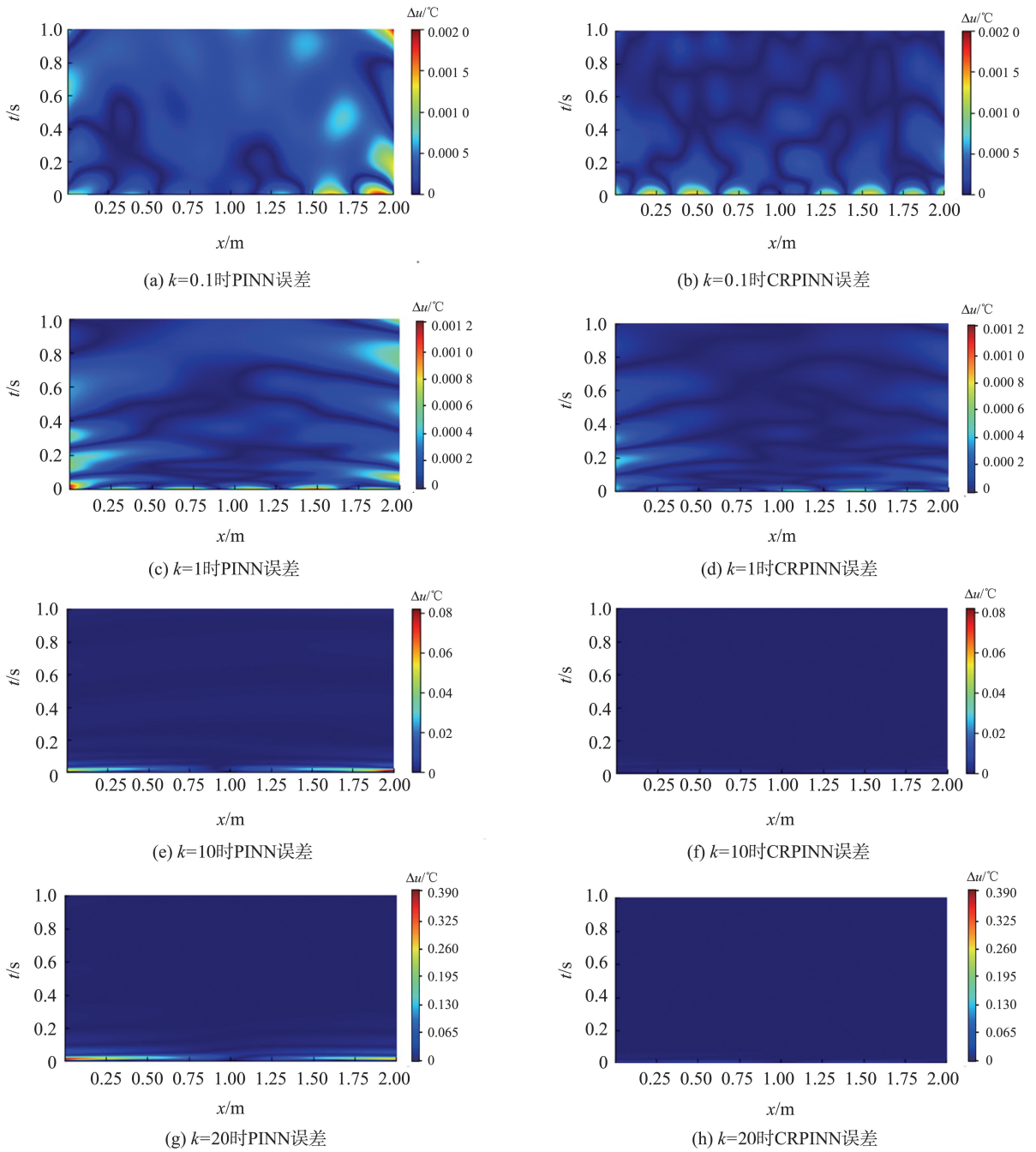


图 12 不同扩散系数下热传导方程预测结果绝对误差的绝对值对比图

Fig.12 Comparison chart of absolute values of predicted absolute errors in heat conduction equation results under different diffusion coefficients

图 13 展示了在 $k=20$ 时热传导方程的解随空间域 x 的变化曲线,可以明显看出 $t=0$ s 时,常规 PINN 和 CRPINN 的结果曲线均与精确解吻合较好,但是随着时间增加,常规 PINN 结果曲线与精确解相差越来越大。与常规 PINN 相比,CRPINN 预测结果曲线与精确解一致性更高。

与 3.1 节中波动方程试验设置类似,同样采用 3 种递进式试验来验证惩罚系数、学习率以及损失函数动态调整策略对结果影响。表 2 给出了不同试验设置下 PINN 预测解 L_2 相对误差。从表 2 中可以推断出与波动方程相同的试验结论,惩罚系数、指数衰减学习率、课程正则化策略均能有效提高预测结果的精度。

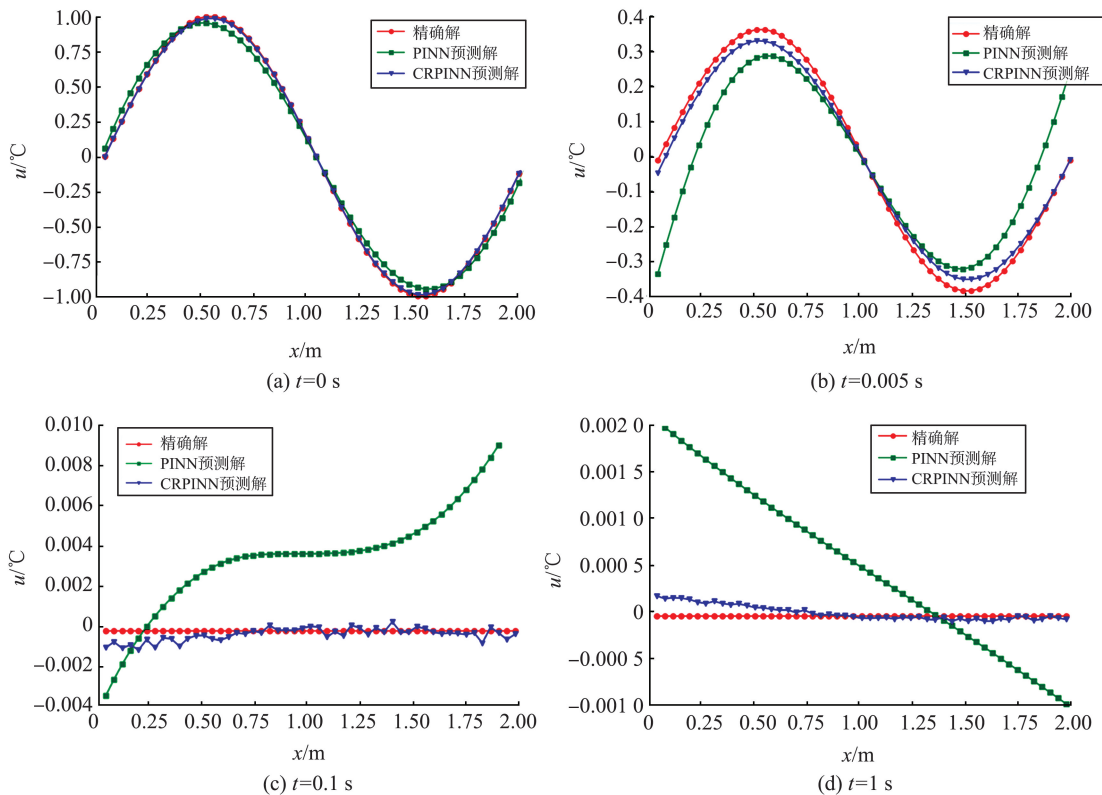


图 13 当 $k=20$ 时热传导方程的解随空间域 x 的变化曲线对比图

Fig. 13 When $k=20$, the solution of the heat conduction equation changes with the space domain x

表 2 热传导方程在不同扩散系数下、不同训练设定下的 L_2 相对误差

Table 2 L_2 relative error of heat conduction equation under different diffusion coefficients and different training Settings

扩散系数 k	L_2 相对误差			
	常规 PINN	PINN+惩罚系数	PINN+惩罚系数+ 指数衰减学习率	PINN+惩罚系数+ 指数衰减学习率+课程正则化
0.1	0.000 760	0.000 481	0.000 412	0.000 399
1	0.000 933	0.000 628	0.000 519	0.000 405
10	0.067 530	0.013 722	0.009 779	0.004 844
20	0.423 511	0.121 265	0.086 353	0.045 025

4 结论

提出了一种基于课程正则化的物理信息神经网络渐进式训练策略,使学习任务从简单逐步过渡到复杂任务,降低了 PINN 优化难度。采用常规 PINN 和优化后 PINN 对不同系数下波动方程和热传导方程进行了求解,验证了本研究所提出方法有效性,主要结论如下。

(1) PINN 的损失景观受损失函数中物理信息部分残差项影响较大,当 PDE 所表征物理信息梯度较大时,损失景观更加复杂, PINN 训练难度随之增大,难以得到预期结果。使用本研究提出的基于课程正则化的损失函数动态调整策略,能够使 PINN 损失景观变得平滑,从而降低训练难度,提高 PINN

的数值稳定性。数值试验表明,与常规 PINN 相比,本研究方法在求解物理信息梯度变化较大的偏微分方程时,能将计算精度提高一个数量级。

(2) 通过惩罚系数调整损失函数中各残差项权重配比,当 $\beta=100$ 时能够有效平衡数据部分和物理信息部分损失,使得测试误差显著下降。

(3) 优化器采取固定步长指数衰减学习率进行优化,能够避免在优化前期陷入局部最小值,并以较高精度接近全局最小值,使得 PINN 预测性能能进一步提高。

本研究所提出的渐进式训练策略可以用于任意一种与时间相关 PDE 的求解,同时可以结合其它 PINN 框架如变分 PINN、基于域分解的并行式 PINN、基于重采样策略的 PINN 等,进一步提高 PINN 的预测精度或计算效率,具有良好扩展性。

模型中引入了两个需要人工设定的超参数:一是算法1中的系数调整步长 α ,步长 α 需要根据不同的任务进行调整以达到精度和效率的平衡,构建自适应步长调整机制解决步长 α 的设定问题;二是式(7)中损失函数惩罚系数 β ,下一步将构建自适应损失函数权重策略来自动优化 β 以达到各项损失的平衡。

参考文献:

- [1] ALTERMAN Z, KARAL F C. Propagation of elastic waves in layered media by finite difference methods[J]. Bulletin of the Seismological Society of America, 1968, 58(1): 367-398.
- [2] KUHLEMEYER R L, LYSMER J. Finite element method accuracy for wave propagation problems[J]. Journal of the Soil Mechanics and Foundations Division, 1973, 99(5): 421-427.
- [3] JASAK H. Error analysis and estimation for the finite volume method with applications to fluid flows[D]. London, UK: Imperial College London, 1996.
- [4] CHENG A H D, CHENG D T. Heritage and early history of the boundary element method [J]. Engineering Analysis with Boundary Elements, 2005, 29(3): 268-302.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [6] LI H. Deep learning for natural language processing: advantages and challenges [J]. National Science Review, 2018, 5(1): 24-26.
- [7] WANG X G, DENG X B, FU Q, et al. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT [J]. IEEE Transactions on Medical Imaging, 2020, 39(8): 2615-2625.
- [8] LI Y, MAO W T, WANG G S, et al. A general-purpose machine learning framework for predicting singular integrals in boundary element method[J]. Engineering Analysis with Boundary Elements, 2020, 117: 41-56.
- [9] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations [J]. Journal of Computational Physics, 2019, 378: 686-707.
- [10] JIANG X C, WANG H, LI Y. A physics-data-driven bayesian method for heat conduction problems [EB/OL]. (2021-09-02) [2023-07-03]. <https://arxiv.org/abs/2109.00996.html>
- [11] WANG H P, LIU Y, WANG S Z. Dense velocity reconstruction from particle image velocimetry/particle tracking velocimetry using a physics-informed neural network[J]. Physics of Fluids, 2022, 34(1): 017116.
- [12] ZENG S J, ZHANG Z, ZOU Q S. Adaptive deep neural networks methods for high-dimensional partial differential equations [J]. Journal of Computational Physics, 2022, 463: 111232.
- [13] DONG S C, YANG J L. Numerical approximation of partial differential equations by a variable projection method with artificial neural networks [J]. Computer Methods in Applied Mechanics and Engineering, 2022, 398: 115284.
- [14] JAGTAP A D, KAWAGUCHI K J, KARNIADAKIS G E. Adaptive activation functions accelerate convergence in deep and physics-informed neural networks[J]. Journal of Computational Physics, 2020, 404: 109136.
- [15] JAGTAP A D, KHARAZMI E, KARNIADAKIS G E. Conservative physics-informed neural networks on discrete domains for conservation laws: applications to forward and inverse problems[J]. Computer Methods in Applied Mechanics and Engineering, 2020, 365: 113028.
- [16] ZOBEIRY N, HUMFELD K D. A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications [J]. Engineering Applications of Artificial Intelligence, 2021, 101: 104232.
- [17] WANG S F, TENG Y J, PERDIKARIS P. Understanding and mitigating gradient flow pathologies in physics-informed neural networks[J]. SIAM Journal on Scientific Computing, 2021, 43(5): 3055-3081.
- [18] MOSELEY B, MARKHAM A, NISSEN-MEYER T. Finite basis physics-informed neural networks (FBPINNs): a scalable domain decomposition approach for solving differential equations[J]. Advances in Computational Mathematics, 2023, 49(4): 62.
- [19] LI J H, CHEN J C, LI B. Gradient-optimized physics-informed neural networks (GOPINNs): a deep learning method for solving the complex modified KdV equation [J]. Nonlinear Dynamics, 2022, 107: 781-792.
- [20] KHARAZMI E, ZHANG Z, KARNIADAKIS G E. Variational physics-informed neural networks for solving partial differential equations [EB/OL]. (2019-11-27) [2023-07-03]. <https://arxiv.org/abs/1912.00873.html>
- [21] RAISSI M, YAZDANI A, KARNIADAKIS G E. Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations[J]. Science, 2020, 367(6481): 1026-1030.
- [22] LU L, MENG X H, MAO Z P, et al. DeepXDE: a deep learning library for solving differential equations

- [J]. *SIAM Review*, 2021, 63(1): 208-228.
- [23] XU K, DARVE E. ADCME: learning spatially-varying physical fields using deep neural networks [EB/OL]. (2020-11-24) [2023-07-03]. <https://arxiv.org/abs/2011.11955.html>.
- [24] HAGHIGHAT E, JUANES R. SciANN: A Keras/TensorFlow wrapper for scientific computations and physics-informed deep learning using artificial neural networks [J]. *Computer Methods in Applied Mechanics and Engineering*, 2021, 373: 113552.
- [25] KRISHNAPRIYAN A, GHOLAMI A, ZHE S, et al. Characterizing possible failure modes in physics-informed neural networks [J]. *Advances in Neural Information Processing Systems*, 2021, 34: 26548-26560.
- [26] EPELBAUM T. Deep learning: technical introduction [EB/OL]. (2017-09-05) [2023-07-03]. <https://arxiv.org/abs/1709.01412.html>
- [27] BAYDIN A G, PEARLMUTTER B A, RADUL A A, et al. Automatic differentiation in machine learning: a survey [J]. *Journal of Machine Learning Research*, 2018, 18: 1-43.
- [28] BENGIO Y, LOURADOUR J, COLLOBERT R. Curriculum learning [C]// *Proceedings of the 26th Annual International Conference on Machine Learning*. New York, USA: Association for Computing Machinery, 2009: 41-48.
- [29] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. (2014-12-22) [2023-07-03]. <https://arxiv.org/abs/1412.6980.html>.
- [30] BERAHAS A S, NOCEDAL J, TAKAC M. A multi-batch *L-BFGS* method for machine learning [J]. *Advances in Neural Information Processing Systems*, 2016, 29: 1055-1063.

(编辑:陈燕)

(上接第10页)

- [28] WANG F F, SUN Z J, GAN Q, et al. Influential node identification by aggregating local structure information [J]. *Physica A: Statistical Mechanics and its Applications*, 2022, 593: 126885.
- [29] QIU L Q, ZHANG J Y, TIAN X B, et al. Ranking influential nodes in complex networks based on local and global structures [J]. *Applied Intelligence*, 2021, 51(7): 1-14.
- [30] 郭茂林, 包崇明, 周丽华, 等. 基于 TOPSIS 的异质网络影响力最大化 [J]. *山东大学学报(工学版)*, 2022, 52(2): 31-40.
- GUO Maolin, BAO Chongming, ZHOU Lihua, et al. Influence maximization of heterogeneous networks based on TOPSIS [J]. *Journal of Shandong University (Engineering Science)*, 2022, 52(2): 31-40.
- [31] BAE J, KIM S. Identifying and ranking influential spreaders in complex networks by neighborhood coreness [J]. *Physica A: Statistical Mechanics & Its Applications*, 2014, 395(4): 549-559.
- [32] 任卓明, 邵凤, 刘建国, 等. 基于度与集聚系数的网络节点重要性度量方法研究 [J]. *物理学报*, 2013, 62(12): 128901.
- REN Zhuoming, SHAO Feng, LIU Jianguo, et al. Node importance measurement based on the degree and clustering coefficient information [J]. *Acta Physica Sinica*, 2013, 62(12): 128901.

(编辑:郭少华)