

## 分数阶非因果BP神经网络模型

黄晶晶, 王建宏

南通大学 理学院, 江苏 南通 226019

**摘 要:** 由一阶因果、反因果微分的定义推导出 Caputo 分数阶因果、反因果微积分, 并在此基础上定义 Caputo 分数阶非因果微积分。将它们分别应用于 BP 神经网络的反向传播过程中对权值进行处理, 产生了 Caputo 分数阶因果、反因果和非因果 BP 神经网络模型。为了方便对比, 将这些模型分别对波士顿房屋数据集和 MNIST 数据集进行处理。模拟结果表明: 在整数阶因果、反因果和非因果的模型之间, 整数阶非因果模型的结果最好; 分数阶因果、反因果和非因果模型分别与其相应的整数阶模型进行比较, 得出分数阶模型得到的准确率比整数阶的高; 在分数阶因果、反因果和非因果的模型之间, 非因果的准确性最高。总的来说, Caputo 分数阶因果、反因果和非因果微积分都对传统 BP 神经网络有优化作用, 尤其是分数阶非因果微积分的优化效果最好。

**关键词:** 因果; 反因果; 非因果; Caputo 分数阶微积分; BP 神经网络

**文献标志码:** A **中图分类号:** TP183 **doi:** 10.3778/j.issn.1002-8331.2012-0072

### Fractional-Order Non-causal BP Neural Networks Model

HUANG Jingjing, WANG Jianhong

School of Science, Nantong University, Nantong, Jiangsu 226019, China

**Abstract:** Based on the definition of the first-order causal and the first-order anti-causal derivative, the anti-causal Caputo derivative is deduced. Then a non-causal Caputo derivative can be given. By applying them to the back propagation process of BP neural network to deal with the weight, the causal, anti-causal and non-causal BP neural network model are generated. For the convenience of comparison, these models are processed for the Boston Housing dataset and the MNIST dataset separately. The results of the simulation show that the integer-order, non-causal model has the best results among the integer-order models; Fractional-order causal, anti-causal and non-causal models are compared with their corresponding integer-order models separately, then the results of the fractional-order models are better. The accuracy of the non-causal is the highest among the fractional-order models. In general, Caputo causal, anti-causal and non-causal calculus all have positive influence on BP neural network, especially the fractional-order non-causal calculus has the best effect.

**Key words:** causal; anti-causal; non-causal; Caputo derivative; BP neural network

BP 神经网络是一种按照误差逆向传播算法训练的多层前馈神经网络, 是应用最广泛的神经网络, 被广泛应用于预测<sup>[1]</sup>、分类模式识别<sup>[2]</sup>等方面。但是, BP 神经网络有着收敛速度慢、耗时较长等缺陷。近年来, 出现了很多对 BP 神经网络进行优化的成果<sup>[3-7]</sup>。特别地, 分数阶的记忆性和遗传性在一定程度上弥补了神经网络的不足。从理论角度出发, 分数阶 BP 算法的误差函数具有单调性和收敛性, 数值模拟的结果也成功验证了这项理论<sup>[5]</sup>。此外, 分数阶 BP 算法已被成功运用于实际应用中, 例如: 对分类数据集进行处理, 能够得到比整数阶 BP 算法更高的分类准确率<sup>[5]</sup>, 且有良好的泛化能力<sup>[6]</sup>; 能

够提高预测的科学性和准确性, 具有更高的预测精度<sup>[7]</sup>。由此可见, 分数阶微积分对 BP 神经网络有很好的优化效果。

分数阶微积分<sup>[8]</sup>这一重要的数学分支诞生在 1695 年, 几乎和经典微积分同时出现。目前, 常见的分数阶微积分定义有 Riemann-Liouville (RL) 定义、Grünwald-Letnikov (GL) 定义、Caputo 定义<sup>[9]</sup>。近一些年分数阶微积分的理论被成功应用到基础研究和工程应用中, 例如图像处理、信号处理<sup>[10]</sup>。随着对分数阶微积分的应用研究的发展, 一些研究者将分数阶微积分理论与因果、反因果、非因果信号处理技术相结合<sup>[11]</sup>, 得出了分数阶因

**基金项目:** 全国统计科学研究项目 (2020LY020); 南通市科技计划项目 (MS12021058)。

**作者简介:** 黄晶晶 (1995—), 女, 硕士研究生, 研究方向为大数据、数学建模及其应用, E-mail: 1061221537@qq.com; 王建宏 (1979—), 通信作者, 男, 博士, 教授, 研究方向为运筹学与控制论、大数据、数学建模及其应用。

**收稿日期:** 2020-12-03 **修回日期:** 2021-01-18 **文章编号:** 1002-8331(2021)23-0091-07

果、反因果、非因果微积分的定义。因果系统为线性时不变系统中一类非常重要的系统,其输出信号由输入信号的现在和过去时刻决定,得到的输出信号为因果信号。相反,反因果系统是指输出信号只与输入信号的现在和将来时刻有关系的系统,其输出输入信号为反因果信号。此外,若输出信号依赖于输入信号的现在、过去和将来时刻的输入,这样的系统称为非因果系统,其输出输入信号为非因果信号。非因果系统常在控制理论和信号处理中被提到<sup>[12-13]</sup>。分数阶非因果微积分实质上是分数阶因果和反因果微积分的加权和。它在保留分数阶的灵活性的情况下,拥有和整数阶导数相同的相频特性<sup>[14]</sup>。因此,从理论上来说分数阶非因果微积分较优。此外,将其用于图像去噪能在有效平滑噪声的同时抑制相位偏移<sup>[15]</sup>;将其用于检波具有较高的检测准确率和定位精度<sup>[14]</sup>。显而易见,分数阶非因果微积分有很高的研究价值。

基于分数阶非因果微积分和分数阶因果BP神经网络的研究,本文构建了分数阶非因果BP神经网络以达到比分数阶因果BP神经网络更好的优化效果。将一阶和Caputo分数阶因果、反因果、非因果微积分分别应用到BP算法的反向传播过程中对权值进行调整,产生了六个不同的模型。为了验证Caputo分数阶非因果BP神经网络的优化效果,需要对这六个模型进行比较。为了方便对比,将这些模型对波士顿房屋数据集和MNIST数据集进行处理。本文的研究需体现非因果和分数阶的优越性,故本文进行了三个比较:整数阶的三个模型之间的对比;分数阶的模型与其相应的整数阶模型之间的对比;分数阶的三个模型之间的对比。

## 1 Caputo 分数阶微积分

本文研究的Caputo非因果微积分实际上是由Caputo因果、反因果微积分线性加权组合成的。

### 1.1 Caputo 分数阶因果微积分

对于函数  $f(t)$ , 其一阶因果导数可写成  $D_t f(t) = \lim_{h \rightarrow 0^+} \frac{f(t) - f(t-h)}{h}$ 。结合Caputo分数阶微积分公式可得到Caputo分数阶因果微积分的定义。

定义1(Caputo分数阶因果微积分)  $\alpha$  阶的Caputo分数阶因果微积分的定义如下所示:

$${}_t \text{Caputo}_a D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_a^t (t-\tau)^{n-\alpha-1} \cdot f^{(n)}(\tau) d\tau \quad (1)$$

其中,  ${}_t \text{Caputo}_a D_t^\alpha$  表示Caputo分数阶因果算子,  $\alpha$  是分数阶,  $[a, t]$  是  $f(t)$  的积分区间。

$\alpha$  在区间  $(0, 1)$  时, 公式就变为:

$${}_t \text{Caputo}_a D_t^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \int_a^t (t-\tau)^{-\alpha} \cdot f'(\tau) d\tau \quad (2)$$

### 1.2 Caputo 分数阶反因果微积分

对于函数  $f(t)$ , 其一阶反因果导数可写成  $D_r f(t) =$

$\lim_{h \rightarrow 0^+} \frac{f(t) - f(t+h)}{h}$ 。由此可推导出Caputo分数阶反因果微积分公式。

定义2(Caputo分数阶反因果微积分)  $\alpha$  阶的Caputo分数阶反因果微积分的定义如下所示:

$${}_r \text{Caputo}_b D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_t^b (t-\tau)^{n-\alpha-1} \cdot f^{(n)}(\tau) d\tau \quad (3)$$

其中,  ${}_r \text{Caputo}_b D_t^\alpha$  表示Caputo分数阶反因果算子,  $[t, b]$  是  $f(t)$  的积分区间。

当  $\alpha$  在区间  $(0, 1)$  时, 公式就变为:

$${}_r \text{Caputo}_b D_t^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \int_t^b (t-\tau)^{-\alpha} f'(\tau) d\tau \quad (4)$$

## 1.3 Caputo 分数阶非因果微积分

本文研究的分数阶非因果微积分, 其实质是分数阶因果和反因果微积分的加权和<sup>[6, 8]</sup>。由此可得出一个相应的Caputo分数阶非因果微积分, 其定义如下所示。

定义3(Caputo分数阶非因果微积分) 当  $\alpha$  在区间  $(0, 1)$  时, 定义Caputo分数阶非因果微积分算子  ${}_a \text{Caputo}_b D_t^\alpha$  为Caputo分数阶因果微积分算子  ${}_t \text{Caputo}_a D_t^\alpha$  与Caputo分数阶反因果微积分算子  ${}_r \text{Caputo}_b D_t^\alpha$  的加权和:

$${}_a \text{Caputo}_b D_t^\alpha f(t) = \frac{1}{2 \sin \frac{\alpha\pi}{2}} ({}_t \text{Caputo}_a D_t^\alpha f(t) - {}_r \text{Caputo}_b D_t^\alpha f(t)) \quad (5)$$

## 2 基于Caputo分数阶微积分的BP算法

在这个部分, 将上面提出的Caputo分数阶因果、反因果、非因果微积分分别代入BP神经网络中, 得到了新的BP算法。

BP神经网络一共有三层: 输入层、隐含层和输出层, 设输入层有  $m$  个, 隐含层有  $n$  个, 输出层有 1 个神经元, 以及输入的样本共用  $J$  个。第  $j$  ( $j=1, 2, \dots, J$ ) 个输入样本和其对应的期望输出分别用  $X^j$  和  $O^j$  表示。输入层和隐含层之间的权重为  $W1 = (\nu_{iq})_{n \times m}$ , 并且隐含层和输出层之间的权重为  $W2 = (\mu_1, \mu_2, \dots, \mu_n)^T$ 。  $\nu_{iq}$  是输入层的第  $q$  ( $q=1, 2, \dots, m$ ) 个神经元和隐含层第  $i$  ( $i=1, 2, \dots, n$ ) 个神经元之间的权重。此外,  $\mu_p$  ( $p=1, 2, \dots, n$ ) 是隐含层第  $p$  个神经元和输出层的那个神经元之间的权重。令  $\nu_i = (\nu_{i1}, \nu_{i2}, \dots, \nu_{im})^T$ 。假设  $g$  和  $f$  是隐含层和输出层的传递函数。

令  $\theta^{ij} = \nu_{i1} x_1^j + \nu_{i2} x_2^j + \dots + \nu_{im} x_m^j = \nu_i \cdot X^j$  是第  $j$  个输入样本对第  $i$  个隐含层神经元的输入, 输出层的输入为  $\xi^j = W2 \cdot G(W1 \cdot X^j)$ , 其实际输出为  $y^j = f(W2 \cdot G(W1 X^j))$ 。对任意一个权重都有一个误差函数:

$$E(W) = \frac{1}{2} \sum_{j=1}^J (O^j - f(W2 \cdot G(W1 \cdot X^j)))^2 = \sum_{j=1}^J f_j(W2 \cdot G(W1 \cdot X^j)) \quad (6)$$

其中,  $f_j(t) = \frac{1}{2}(O^j - f(t))^2$ ,  $t \in \mathbb{R}$ ,  $f_j(\cdot)$  是函数  $G$  和  $f$  构成的组合函数。已知  $W^0 = (W_1^0, W_2^0)$  是初始权重, 第  $k$  个权重为  $W^k = (W_1^k, W_2^k)$ 。本文所研究的分数阶BP算法对权值迭代更新的过程做了改变: 误差函数对每个权值做Caputo分数阶偏导数。

## 2.1 基于Caputo分数阶因果微积分的BP算法

$$\mu_i^{k+1} = \mu_i^k - \eta \cdot {}_l\text{Caputo}_{cmin} D_{\mu_i^k}^\alpha E(W)$$

$$\nu_{ir}^{k+1} = \nu_{ir}^k - \eta \cdot {}_l\text{Caputo}_{cmin} D_{\nu_{ir}^k}^\alpha E(W)$$

其中,  $\eta > 0$  是学习速率,  $0 < \alpha < 1$  是分数阶,  $cmin = \min(\mu_i^k, \nu_{ir}^k)$  ( $k \in \mathbb{N}, i = 1, 2, \dots, n, r = 1, 2, \dots, m$ )。

$h(s(t))$  是一个组合函数, 根据Caputo分数阶导数的定义,  $h(s(t))$  关于  $t$  的  $\alpha$  阶导数为:

$${}_l\text{Caputo}_a D_t^\alpha h(s) = \frac{\partial}{\partial s}(h(s)) \cdot {}_l\text{Caputo}_a D_t^\alpha s(t) \quad (7)$$

根据公式(7), 可以算出  ${}_l\text{Caputo}_{cmin} D_{\mu_i^k}^\alpha E(W)$  和  ${}_l\text{Caputo}_{cmin} D_{\nu_{ir}^k}^\alpha E(W)$ 。

第一部分:  ${}_l\text{Caputo}_{cmin} D_{\mu_i^k}^\alpha E(W)$

$${}_l\text{Caputo}_{cmin} D_{\mu_i^k}^\alpha E(W) = \frac{\partial E(W)}{\partial \zeta^j} \cdot {}_l\text{Caputo}_{cmin} D_{\mu_i^k}^\alpha (\zeta^j) \quad (8)$$

其中:

$$\frac{\partial E(W)}{\partial \zeta^j} = \sum_{j=1}^J f_j'(W_2 \cdot G(W_1 \cdot X^j))$$

根据公式(2), 当  $0 < \alpha < 1$  时, 能够得到:

$$\begin{aligned} {}_l\text{Caputo}_{cmin} D_{\mu_i^k}^\alpha (\zeta^j) &= \frac{1}{\Gamma(1-\alpha)} \int_{cmin}^{\mu_i^k} (\mu_i^k - \tau)^{-\alpha} \cdot \\ &g(\nu_i^k \cdot X^j) d\tau = \frac{1}{\Gamma(1-\alpha)} g(\nu_i^k \cdot X^j) \int_{cmin}^{\mu_i^k} (\mu_i^k - \tau)^{-\alpha} d\tau = \\ &\frac{1}{\Gamma(1-\alpha)} \frac{1}{1-\alpha} g(\nu_i^k \cdot X^j) (\mu_i^k - cmin)^{1-\alpha} \end{aligned} \quad (9)$$

其中,  $k \in \mathbb{N}$ ,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, J$ 。则误差关于  $\mu_i^k$  的具体公式为:

$$\begin{aligned} {}_l\text{Caputo}_{cmin} D_{\mu_i^k}^\alpha E(W) &= \frac{1}{(1-\alpha)\Gamma(1-\alpha)} \cdot \sum_{j=1}^J f_j'(W_2 \cdot \\ &G(W_1 \cdot X^j)) g(\nu_i^k \cdot X^j) (\mu_i^k - cmin)^{1-\alpha} \end{aligned} \quad (10)$$

第二部分:  ${}_l\text{Caputo}_{cmin} D_{\nu_{ir}^k}^\alpha E(W)$

类似于公式(10)有:

$${}_l\text{Caputo}_{cmin} D_{\nu_{ir}^k}^\alpha E(W) = \frac{\partial E(W)}{\partial \theta^{i,j}} \cdot {}_l\text{Caputo}_{cmin} D_{\nu_{ir}^k}^\alpha (\theta^{i,j}) \quad (11)$$

其中:

$$\frac{\partial E(W)}{\partial \theta^{i,j}} = \sum_{j=1}^J f_j'(W_2 \cdot G(W_1 X^j)) \mu_i^k g'(\nu_i \cdot X^j)$$

则  $\theta^{i,j}$  关于  $\nu_{ir}^k$  的Caputo分数阶因果微积分:

$$\begin{aligned} {}_l\text{Caputo}_{cmin} D_{\nu_{ir}^k}^\alpha (\theta^{i,j}) &= \frac{1}{\Gamma(1-\alpha)} \int_{cmin}^{\nu_{ir}^k} (\nu_{ir}^k - \tau)^{-\alpha} x_r^j d\tau = \\ &\frac{1}{\Gamma(1-\alpha)} x_r^j \frac{1}{1-\alpha} (\nu_{ir}^k - \tau)^{-\alpha} \Big|_{cmin}^{\nu_{ir}^k} = \\ &\frac{1}{(1-\alpha)\Gamma(1-\alpha)} x_r^j (\nu_{ir}^k - cmin)^{1-\alpha} \end{aligned} \quad (12)$$

由上式可以得出:

$$\begin{aligned} {}_l\text{Caputo}_{cmin} D_{\nu_{ir}^k}^\alpha E(W) &= \frac{1}{(1-\alpha)\Gamma(1-\alpha)} \cdot \sum_{j=1}^J f_j'(W_2 \cdot \\ &G(W_1 X^j)) \mu_i^k g'(\nu_i \cdot X^j) x_r^j (\nu_{ir}^k - cmin)^{1-\alpha} \end{aligned} \quad (13)$$

## 2.2 基于Caputo分数阶反因果微积分的BP算法

$$\mu_i^{k+1} = \mu_i^k - \eta \cdot {}_r\text{Caputo}_{cmax} D_{\mu_i^k}^\alpha E(W)$$

$$\nu_{ir}^{k+1} = \nu_{ir}^k - \eta \cdot {}_r\text{Caputo}_{cmax} D_{\nu_{ir}^k}^\alpha E(W)$$

依据公式(10)和(13)的计算方法, 可以算出  ${}_r\text{Caputo}_{cmax} D_{\mu_i^k}^\alpha E(W)$  和  ${}_r\text{Caputo}_{cmax} D_{\nu_{ir}^k}^\alpha E(W)$ , 具体公式如下:

$$\begin{aligned} {}_r\text{Caputo}_{cmax} D_{\mu_i^k}^\alpha E(W) &= \frac{1}{(1-\alpha)\Gamma(1-\alpha)} \cdot \sum_{j=1}^J f_j'(W_2 \cdot \\ &G(W_1 \cdot X^j)) g(\nu_i^k \cdot X^j) (cmax - \mu_i^k)^{1-\alpha} \end{aligned} \quad (14)$$

$$\begin{aligned} {}_r\text{Caputo}_{cmax} D_{\nu_{ir}^k}^\alpha E(W) &= \frac{1}{(1-\alpha)\Gamma(1-\alpha)} \cdot \sum_{j=1}^J f_j'(W_2 \cdot \\ &G(W_1 X^j)) \mu_i^k g'(\nu_i \cdot X^j) x_r^j (cmax - \nu_{ir}^k)^{1-\alpha} \end{aligned} \quad (15)$$

## 2.3 基于Caputo分数阶非因果微积分的BP算法

根据Caputo分数阶非因果微积分的定义, 可以得出  ${}_d\text{Caputo} D_{\mu_i^k}^\alpha E(W)$  以及  ${}_d\text{Caputo} D_{\nu_{ir}^k}^\alpha E(W)$ 。最终结果如下。

第一部分:  ${}_d\text{Caputo} D_{\mu_i^k}^\alpha E(W)$

$$\begin{aligned} {}_d\text{Caputo} D_{\mu_i^k}^\alpha E(W) &= \frac{1}{2 \sin \frac{\alpha\pi}{2}} ({}_l\text{Caputo}_{cmin} D_{\mu_i^k}^\alpha E(W) - \\ &{}_r\text{Caputo}_{cmax} D_{\mu_i^k}^\alpha E(W)) \end{aligned} \quad (16)$$

第二部分:  ${}_d\text{Caputo} D_{\nu_{ir}^k}^\alpha E(W)$

$$\begin{aligned} {}_d\text{Caputo} D_{\nu_{ir}^k}^\alpha E(W) &= \frac{1}{2 \sin \frac{\alpha\pi}{2}} ({}_l\text{Caputo}_{cmin} D_{\nu_{ir}^k}^\alpha E(W) - \\ &{}_r\text{Caputo}_{cmax} D_{\nu_{ir}^k}^\alpha E(W)) \end{aligned} \quad (17)$$

## 3 实例分析

为了验证Caputo分数阶非因果BP神经网络模型的效果。将这些模型分别应用到波士顿房屋数据集和MNIST数据集中。通过所得准确率的高低来验证Caputo分数阶非因果的模型的准确率最高。为了保证实验结果的可靠性, 选取了多个学习速率, 隐含层神经元的数目和分数阶进行比较, 最后选取比较好的数值代入这些模型中进行总的比较分析。在这个过程中发现无论是因果、反因果还是非因果的模型, Caputo分数阶BP算法都比一阶的结果更好。此外, 分数阶非因果的结果是这些模型中最好的。整个模拟通过matlab软件完成。

### 3.1 波士顿房屋数据集

波士顿房屋数据集是美国人口普查局在1978年收集的美国马萨诸塞州波士顿住房价格的有关信息, 涵盖



了506条波士顿不同郊区的房屋数据,每条数据14个字段,包含自住房的平均房价(MEDV)和13个影响因素:城镇人均犯罪率(CRIM)、城镇非零售商用土地的比例(INDUS)、住宅平均房间数(RM)等。本文以这13个影响因素为输入,以MEDV为输出。

在这次模拟中,将这506条数据随机排列后,取前404条为波士顿房屋训练集,剩下的102条数据为波士顿房屋测试集。BP神经网络有一个隐含层,隐含层神经元的数量取  $HidNodes=10$ ,其网络结构为  $13 \times 10 \times 1$ 。本文选取的阶数为  $\alpha=1/9, 2/9, \dots, 8/9$  和1,  $\alpha=1$  代表了整数阶的情况。最大迭代次数为1000。每种网络都要训练五次,取平均值得到最终的结果。为了直观地比较不同模型的预测值的可靠性,本节用绝对系数  $R^2$  来显示预测值的拟合效果。以第  $j$  个样本为例,令其预测值为  $y_j$ ,期望输出为  $O^j$ ,则其绝对系数  $R^2$  为:

$$R^2 = 1 - \frac{\sum_{i=1}^J (O^i - y_i)^2}{\sum_{i=1}^J (O^i - \bar{O})^2}$$

其中,  $J$  为数据集中数据组的总组数,  $R^2$  的取值在  $[0, 1]$  内,且值越大预测值的拟合优度越高。经过反复实验,将学习速率定为0.08进行最终的模拟。训练集和测试集经过模拟得到的绝对系数如表1和表2所示。

表1 波士顿房屋测试集的绝对系数

Table 1 Absolute coefficient of Boston housing testing dataset

| 阶数  | 因果    | 反因果   | 非因果   |
|-----|-------|-------|-------|
| 1/9 | 0.851 | 0.826 | 0.935 |
| 2/9 | 0.844 | 0.815 | 0.919 |
| 1/3 | 0.849 | 0.822 | 0.917 |
| 4/9 | 0.869 | 0.828 | 0.921 |
| 5/9 | 0.869 | 0.821 | 0.922 |
| 2/3 | 0.861 | 0.834 | 0.926 |
| 7/9 | 0.861 | 0.831 | 0.911 |
| 8/9 | 0.866 | 0.820 | 0.905 |
| 1   | 0.858 | 0.819 | 0.911 |

表2 波士顿房屋训练集的绝对系数

Table 2 Absolute coefficient of Boston housing training dataset

| 阶数  | 因果    | 反因果   | 非因果   |
|-----|-------|-------|-------|
| 1/9 | 0.897 | 0.824 | 0.934 |
| 2/9 | 0.902 | 0.819 | 0.933 |
| 1/3 | 0.886 | 0.821 | 0.925 |
| 4/9 | 0.893 | 0.825 | 0.915 |
| 5/9 | 0.889 | 0.825 | 0.898 |
| 2/3 | 0.884 | 0.841 | 0.898 |
| 7/9 | 0.880 | 0.841 | 0.890 |
| 8/9 | 0.883 | 0.835 | 0.892 |
| 1   | 0.875 | 0.829 | 0.895 |

由表1和表2可知,无论是整数阶还是分数阶的模型,非因果BP神经网络的预测值的拟合优度最高;对于因果、反因果和非因果BP预测模型,总能找到一个或多个分数阶的预测结果比1阶的好。总的来说,对于波士顿房屋数据集,分数阶非因果BP神经网络的预测值的可靠性最高。

### 3.2 MNIST数据集

MNIST数据集来自美国国家标准与技术研究所,是机器学习领域中非常经典的一个数据集。其中的每张图片由  $28 \times 28$  个像素点构成,也可展开为一个  $784 \times 1$  的一维行向量,里面的每个元素都是  $(0, 255)$  里的数,代表了每个像素的灰度等级。该数据集包含了60000个用于训练的示例和10000个用于测试的示例。在这次模拟中,这60000个训练示例组成了测试集,10000个测试示例组成了测试集<sup>[15]</sup>。

BP神经网络有一个隐含层,隐含层神经元的数量取  $HidNodes=50, 100, 150, 200$ ,其网络结构为  $784 \times \{50, 100, 150, 200\} \times 10$ 。本文选取的阶数依然为  $\alpha=1/9, 2/9, \dots, 8/9$  和1。网络中的学习速率是一个重要的参数,选取  $\eta=1.5, 2, 3, 4$  进行比较选择。最大迭代次数为5。每种网络都要训练5次,取平均值得到最终的训练准确率和测试准确率。

#### 3.2.1 整数阶BP算法的比较分析

本小节分别用1阶因果、反因果和非因果BP算法对训练集和测试集进行处理,结果如图1所示。图1显示在同一个学习速率下,非因果的曲线一直是最高的。可见整数阶非因果模型比整数阶因果、反因果的更好。

#### 3.2.2 分数阶BP算法的比较分析

因果BP神经网络的模拟结果如图2所示:在相同的  $HidNodes$  值下,从每个子图的最高点的准确率可知:测试集和训练集的准确率都随着学习速率的增大而提高;当  $\alpha=8/9$  时准确率最高;当隐含层神经元数  $HidNode$  从50到100时准确率变高,但到150之后就变低。总之,对于因果BP神经网络,分数阶的模型比整数阶的好,且较合适的参数为  $\alpha=8/9$ ,  $HidNodes=100$ ,  $\eta=4$ 。由图3可知,反因果模型的最高准确率基本出现在  $\alpha=8/9$  的时候,且其较合适的  $HidNodes$  值和学习速率与因果模型的相同。

从图4可见,非因果BP神经网络模型的图像与因果、反因果的基本相反:当  $\alpha=8/9$  时,非因果的准确率最低;当  $\alpha=2/9$  时,准确率最高。此外,对于非因果模型较合适的  $HidNodes$  值为150,学习速率为3。

总的来说,无论是因果、反因果还是非因果模型,分数阶的模拟结果比其对应的整数阶的好。

#### 3.2.3 分数阶与传统BP算法的比较分析

已知因果和反因果BP神经网络算法的较合适的参数为  $\eta=4$ ,  $HidNodes=100$ ,  $\alpha=8/9$ 。传统BP神经网络

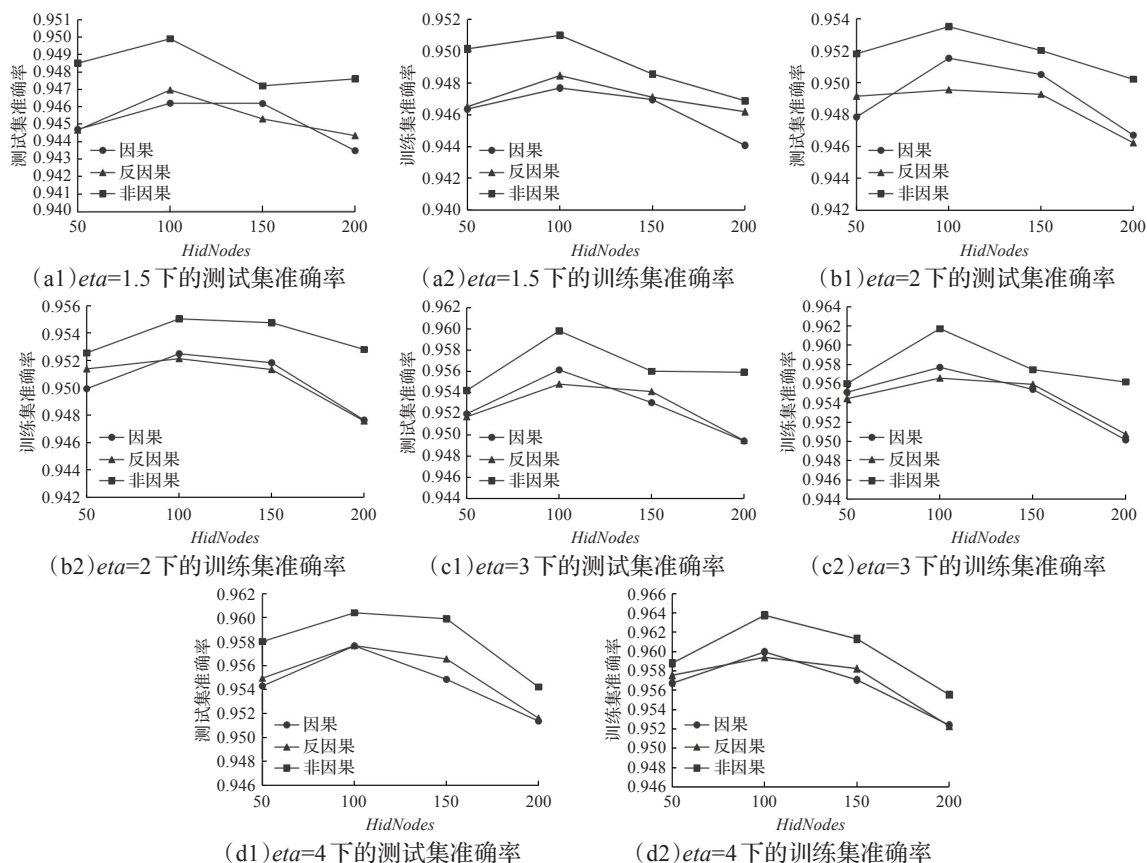


图1 整数阶BP神经网络结果对比图

Fig.1 Comparison between integer-order BP algorithms

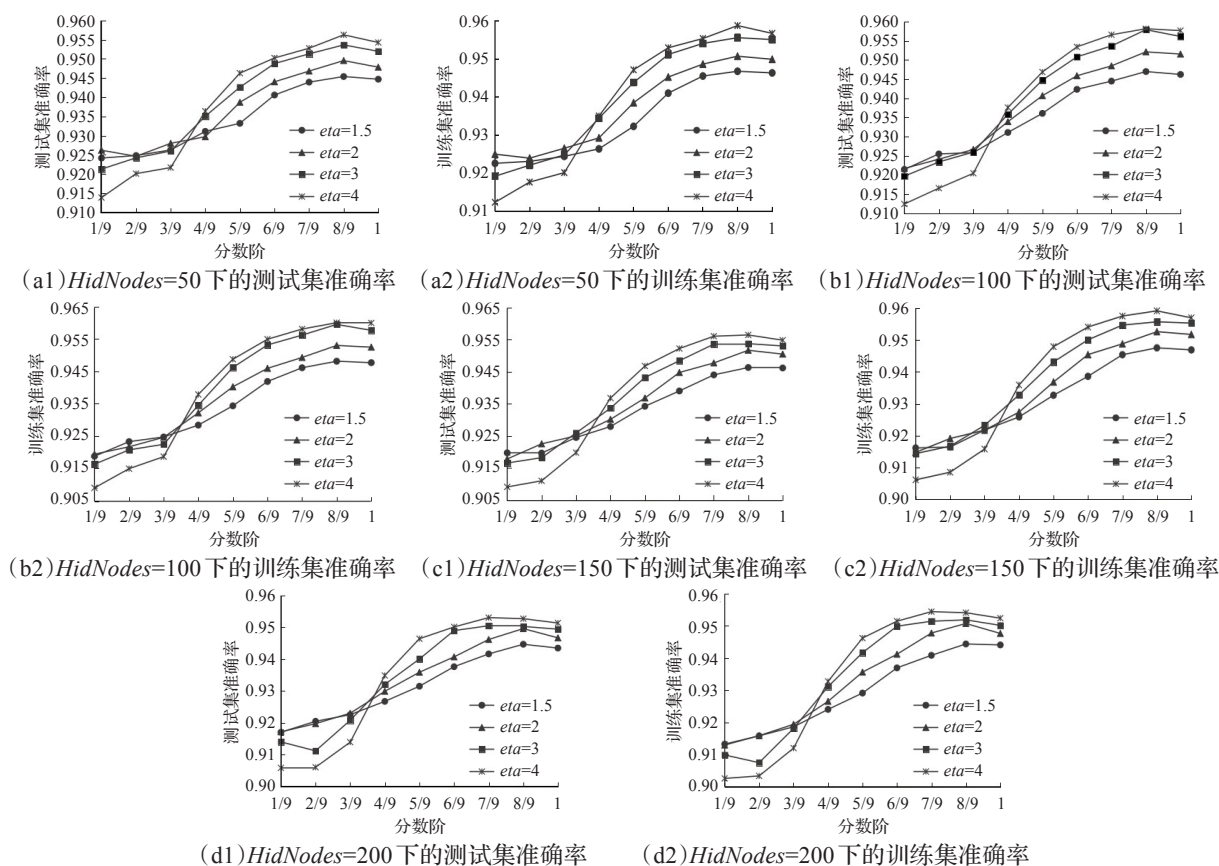


图2 因果BP神经网络结果对比图

Fig.2 Comparison between causal BP algorithms

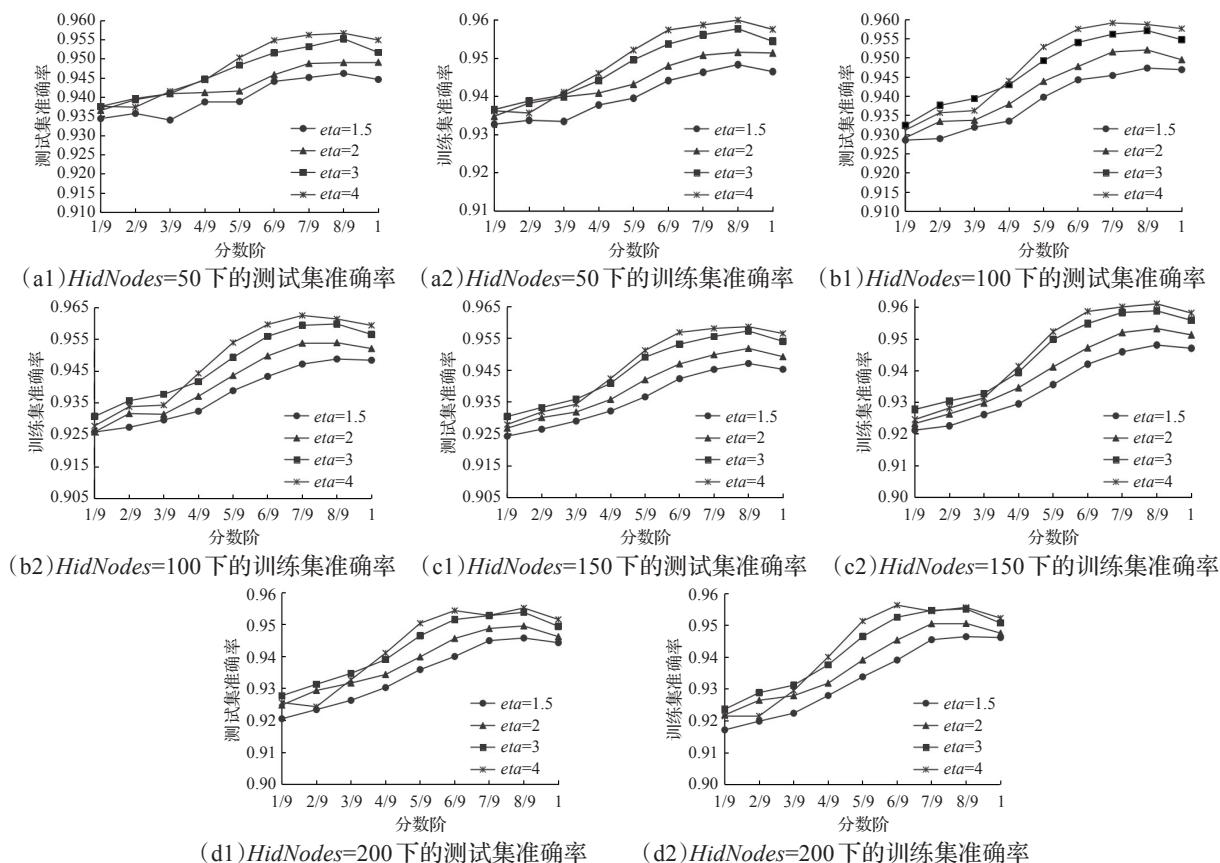


图3 反因果BP神经网络结果对比图

Fig.3 Comparison between anti-causal BP algorithms

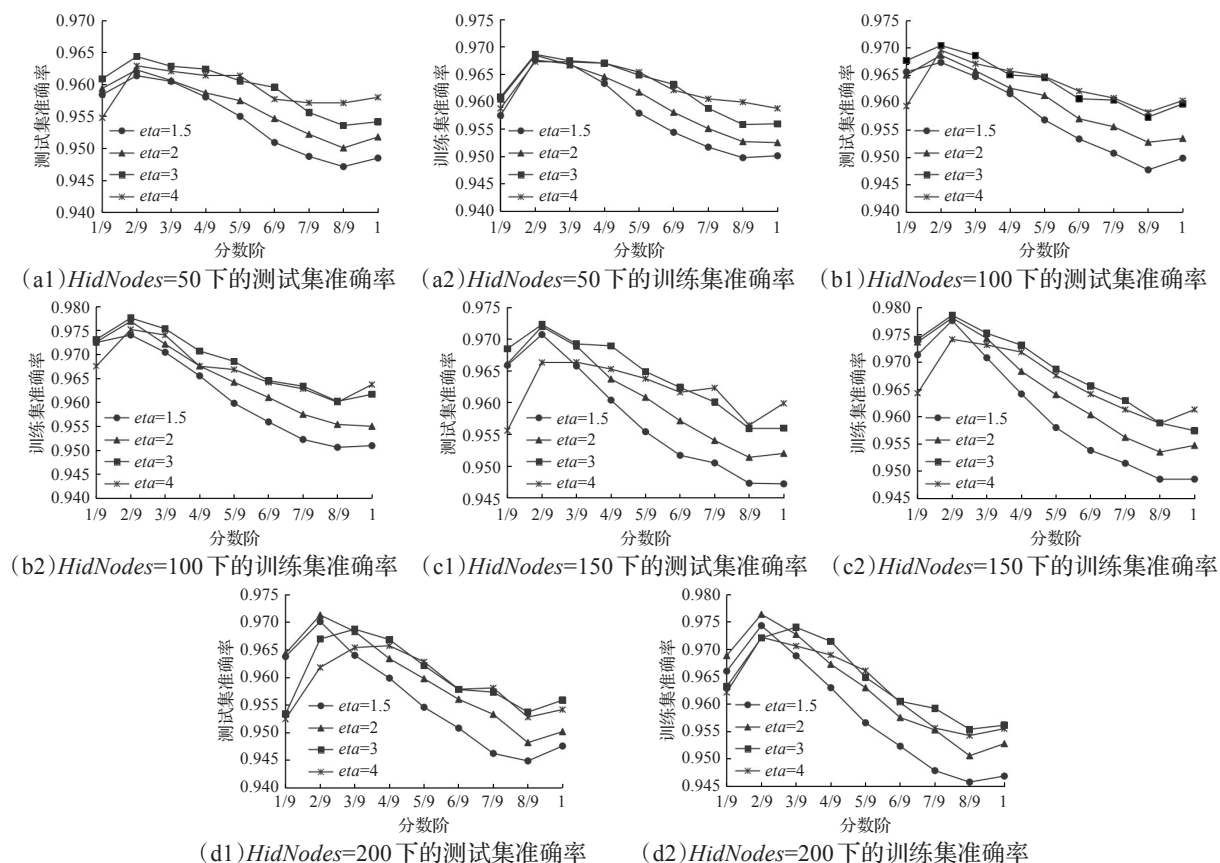


图4 非因果BP神经网络结果对比图

Fig.4 Comparison between non-causal BP algorithms



络和分数阶因果、反因果、非因果BP神经网络在这些取值下的模拟结果如图5所示。由图5可知,分数阶模型的曲线都比传统的高,尤其是分数阶非因果模型。这证明了分数阶因果、反因果和非因果微积分都对BP神经网络有优化作用,其中分数阶非因果微积分的优化作用最好。

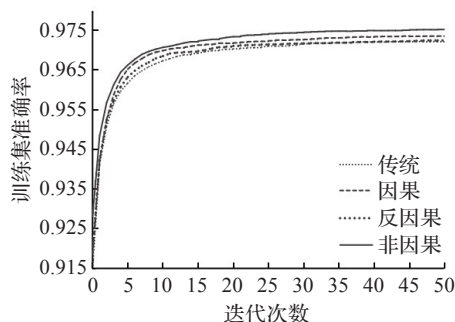


图5 分数阶与传统BP神经网络结果对比图

Fig.5 Comparison between traditional and fractional-order BP algorithms

#### 4 结语

本文在分数阶对BP神经网络具有优化效果的基础上,进一步提出了分数阶非因果微积分对BP神经网络进行优化的想法。结合1阶因果和反因果的概念推导出了反因果Caputo分数阶微积分,并由分数阶非因果微积分的定义得出Caouto非因果微积分。此外,将一阶和Caputo分数阶因果、反因果、非因果微积分分别应用到BP神经网络的反向传播过程中,从而产生了新的BP算法。为了验证分数阶非因果模型是这些模型中准确性最高的,分别用这些模型对波士顿房屋数据集和MNIST数据集进行模拟。模拟的结果和本文预想的一样:分数阶的结果比整数阶的好;非因果模型的预测拟合度和分类准确率都比因果和反因果的高。总之,分数阶非因果BP算法对于预测和分类识别都有更高的准确性。本文验证了分数阶非因果微积分与BP神经网络相结合的可行性,在神经网络优化的研究上有很好的前景。

#### 参考文献:

- [1] 刘承伟,赵洪凯,严豪,等.基于分数阶GM(1,1)与BP神经网络的电力负荷预测[J].数学的实践与认识,2018,48(23):145-151.  
LIU C W,ZHAO H K,YAN H,et al.Power load forecasting based on fractional GM(1,1) and BP neural network[J].Mathematics in Practice and Theory,2018,48(23):145-151.
- [2] 易云清,吕乐群,卢圆圆,等.基于神经网络的复杂信号样式调制识别技术[J].电子信息对抗技术,2020,35(6):16-21.  
YI Y Q,LY L Q,LU Y Y,et al.The modulation recognition technology of complex signal pattern based on neural network[J].Electronic Warfare Technology,2020,35(6):16-21.
- [3] 何明慧,徐怡,王冉,等.改进的粒子群算法优化神经网络

- 及应用[J].计算机工程与应用,2018,54(19):107-113.
- HE M H,XU Y,WANG R,et al.Improved particle swarm optimization neural network and its application[J].Computer Engineering and Applications,2018,54(19):107-113.
- [4] 罗宇卓.基于FPSO优化的BP神经网络算法研究与应用[D].银川:宁夏大学,2018.  
LUO Y Z.Research on BP neural network algorithm and application based on FPSO optimization[D].Yinchuan: Ningxia University,2018.
- [5] WANG J,WEN Y,GOU Y,et al.Fractional-order gradient descent learning of BP neural networks with Caputo derivative[J].Neural Networks,2017,89(89):19-30.
- [6] 温艳青.分数阶神经网络学习算法的设计与理论分析[D].青岛:中国石油大学,2017.  
WEN Y Q.Algorithm design and convergence analysis for fractional-order neural network[D].Qingdao: China University of Petroleum,2017.
- [7] 付华,王福娇,陈子春.基于分数阶神经网络的瓦斯涌出量预测[J].传感器与微系统,2013,32(5):31-34.  
FU H,WANG F J,CHEN Z C.Prediction of gas emission based on fractional order neural network[J].Transducer and Microsystem Technologies,2013,32(5):31-34.
- [8] DAS S.Fractional calculus[M].Berlin:Springer-Verlag,2011.
- [9] SAMKO S,KILBAS A,MARICHEV O.Fractional integrals and derivatives: theory and applications[M].New York: Gordon and Breach Science Publishers,1993.
- [10] 吴媛媛,潘祥,姜太平,等.基于非因果信号处理的分数阶次低通滤波器[J].计算机应用与软件,2017,34(9):192-198.  
WU Y Y,PAN X,JIANG T P,et al.Non-causal signal processing based fractional low-pass Filter[J].Computer Applications and Software,2017,34(9):192-198.
- [11] 刘双双.基于非因果信号处理的分数阶次图像去噪算法研究[D].马鞍山:安徽工业大学,2017.  
LIU S S.Non-causal signal processing based fractional image denoising algorithm[D].Ma' anshan: Anhui University of Technology,2017.
- [12] TSAI M,LIN M,YAU H.Development of command-based iterative learning control algorithm with consideration of friction,disturbance,and noise effects[J].IEEE Transactions on Control Systems Technology,2006,14(3):511-518.
- [13] 朱洪俊.心电信号零相位数字滤波[J].北京生物医学工程,2003(4):260-262.  
ZHU H J.ECG signal processing with zero-phase digital filtering[J].Beijing Biomedical Engineering,2003(4):260-262.
- [14] WANG J,YE Y,GAO X.Fractional 90° phase-shift filtering based on the double-sided Grünwald-Letnikov differintegrator[J].IET Signal Processing,2015,9(4):328-334.
- [15] ANDREI V.Neural network for low-memory IoT devices and MNIST image recognition using kernels based on logistic map[J].Electronics,2020,9(9):1432.