

# V 正交基网络

熊刚强<sup>1</sup> 齐东旭<sup>2,3</sup>

(广东医学院信息工程学院 东莞 523808)<sup>1</sup> (中山大学信息科学与技术学院 广州 510275)<sup>2</sup>  
(澳门科技大学资讯科技学院 澳门)<sup>3</sup>

**摘 要** 为了改进 BP 网络的收敛速度与连续正交基网络无法逼近非连续函数的问题,构造了一类基于 V 正交基的前馈神经网络(简称 V 正交基网络),并研究其收敛性条件与伪逆规则。由于 V 系统是  $L^2([0,1])$  上的一类完备的正交函数系,且 Fourier-V 级数有较快的收敛速度,因此,V 正交基网络有较快的收敛速度,且能有效地逼近一类强间断的一元函数。最后,通过仿真实验证明,V 正交基网络的收敛速度明显优于传统的 BP 网络、小波网络与 Legendre 网络,特别是逼近一类间断点在二进制有理数处的函数时,其优势更加明显。

**关键词** V 系统, BP 神经网络, 小波神经网络, Legendre 网络, 函数逼近

**中图法分类号** TP183 **文献标识码** A

## V Orthonormal Basis Neural Network

XIONG Gang-qiang<sup>1</sup> QI Dong-xu<sup>2,3</sup>

(School of Information Engineering, Guangdong Medical College, Dongguan 523808, China)<sup>1</sup>

(School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510275, China)<sup>2</sup>

(Faculty of Information Technology, Macao University of Science and Technology, Macao, China)<sup>3</sup>

**Abstract** In order to solve the problem that the convergence rate of BP network is not fast, and the neural networks with continuous orthogonal basis cannot approximate discontinuous functions, this paper constructed a class of feed-forward neural networks with V orthonormal basis (referred to as V orthogonal network), and investigated its convergence condition and pseudo-inverse rule. For V system is a class of complete orthonormal systems in  $L^2([0,1])$ , and the convergence rate of Fourier-V series is comparatively fast, the convergence rate of V orthogonal network is also fast, and it can effectively approximate a class of discontinuous functions of one variable. The simulation results also show that the convergence rate of V orthogonal network is obviously faster than that of BP network, wavelet network and Legendre network; if using V orthogonal network to approximate the functions whose breakpoints only appear at dyadic rational, its performance of function approximation becomes much better.

**Keywords** V system, BP network, Wavelet network, Legendre network, Function approximation

## 1 引言

基于误差反向传播(Back Propagation, BP)算法的前馈神经网络被广泛应用到各个科学与工程领域中,但是 BP 神经网络的收敛速度慢且易陷入局部极小值点<sup>[1]</sup>。就这个问题,很多学者提出了各种解决方案,主要思路是改进 BP 算法本身,如:使用启发式信息技术与数值优化技术<sup>[1]</sup>。另一种做法是改进网络结构,用线性无关的基函数作为 BP 网络的隐层神经元,这样可以改善 BP 网络收敛速度,如小波网络<sup>[2,3]</sup>、正交基网络<sup>[4-7]</sup>,但是,我们在实验中发现,这类神经网络无法逼近间断函数,收敛速度很不理想。而这类神经网络有一个明显的特点,即使用连续的基函数作为隐层神经元。一个很自然的想法是:使用非连续的基函数作为隐层神经元去逼近非连续的函数,也许可以获得更加理想的逼近效果。其中常见

的非连续的基函数有 Walsh 函数与 Haar 函数,用小波的观点来看, Walsh 系与 Haar 系的逼近阶为 1,它们本身的收敛速度慢,因此,用它们构造正交基神经网络当然不是最佳选择。本文将使用高次 Haar 系(即 V 系统)<sup>[8]</sup>的基函数作为隐层神经元,构造一类新的 BP 神经网络,并通过仿真实验,发现本文所研究的正交基神经网络在逼近一元函数时可得到非常理想的逼近效果。

## 2 V 系统

### 2.1 V 系统的构造

V 系统由 U 系统的函数生成元经过压缩与平移得到,其中 U 系统的函数生成元是由  $[0,1]$  区间上的 Legendre 多项式构造得到的<sup>[9]</sup>。考虑到连贯性,首先简要地介绍 U 系统的函数生成元的定义。记  $[0,1]$  区间上前  $r$  个 Legendre 多项式为

到稿日期:2010-11-30 返修日期:2011-04-25 本文受国家自然科学基金重点项目(10631080),澳门科学发展基金项目(045/2006/A)资助。

熊刚强(1968—),男,博士,讲师,主要研究方向为图像处理与小波分析;齐东旭(1940—),男,教授,主要研究方向为计算数学、计算机图形学、信息安全。

$v_0^{(0)}, v_1^{(0)}, \dots, v_{r-1}^{(0)}$ , 那么 U 系统的函数生成元  $v_0^{(1)}, v_1^{(1)}, \dots, v_{r-1}^{(1)}$  应满足下面 3 个条件。

- 1) 表达式:  $v_k^{(1)}(x)$  是以  $x = \frac{1}{2}$  为分界点的  $r-1$  次分段多项式,  $k=0, \dots, r-1$ ;
- 2) 正交性:  $\langle v_k^{(i)}, v_l^{(j)} \rangle = \delta(i-j)\delta(k-l)$ , 其中,  $k, l \in \{0, \dots, r-1\}, i, j \in \{0, 1\}, \delta(n) = \begin{cases} 1, & n=0 \\ 0, & n \neq 0 \end{cases}$ ,  $\langle \cdot, \cdot \rangle$  表示内积;
- 3) 对称性:  $v_k^{(1)}(1-x) = (-1)^{r+k} v_k^{(1)}(x)$ , 其中,  $k \in \{0, 1, \dots, r-1\}, x \in [0, 1]$ 。

由定义可以构造  $r$  个函数生成元, 但并不唯一, 若满足文献[9]定理 1 的条件, 则可以构造一类特殊性质的函数生成元。把函数生成元压缩后, 再进行正复制/反复制即可得到 U 系统的其它基; 若把函数生成元压缩后, 再进行平移即可得到 V 系统的其它基, 即

$$v_{l,k}^{(j)}(x) = \sqrt{2^{j-1}} v_l^{(1)}(2^{j-1}x - k)$$

式中,  $k=0, 1, \dots, 2^{j-1}-1, l=0, 1, \dots, r-1$

记  $v_{l,k}^{(j)}$  为 V 系统的第  $j(j>1)$  层基, 它的第 0 层基就是 Legendre 多项式, 第 1 层基就是函数生成元。如果把所得到的基函数按层、生成元的序号、平移量从小到大顺序排列后, 再用统一的符号  $v_n, n=0, 1, \dots$  表示, 这样 V 系统可以表示为

$$\{v_n(x): n=0, 1, \dots\} \quad (1)$$

## 2.2 V 系统的性质

记  $f \sim \sum_{n=0}^{+\infty} a_n v_n$  为函数  $f$  的 Fourier-V 级数, 其中  $a_n = \langle f, v_n \rangle = \int_0^1 f(x) v_n(x) dx$  是 Fourier-V 系数, 称  $S_f(n) = \sum_{n=0}^{n-1} a_n v_n$  为函数  $f$  的 Fourier-V 级数的前  $n$  项的部分和。实际上, U 系统的基可以表示为 V 系统的基与 Walsh 矩阵或倒 Walsh 矩阵的乘积<sup>[10]</sup>, 因此, 函数集(1)是  $L^2([0, 1])$  上一组完备的标准正交基, 即有下面命题成立<sup>[8]</sup>。

**定理 1** 若  $f \in L^2([0, 1])$ , 则  $\lim_{n \rightarrow \infty} \|f - S_f(n)\|_{L^2} = 0$ ; 若  $f \in C([0, 1])$ , 则  $\lim_{n \rightarrow \infty} \|f - S_f(2^n)\|_{L^\infty} = 0$ 。

本质上,  $r-1$  次 V 系统是一类  $L^2([0, 1])$  上分段多项式多小波<sup>[10]</sup>, 比如 0 次 V 系统(即  $r=1$ )就是 Haar 系, 也即  $[0, 1]$  区间上的 Haar 小波, 并且是正交、对称、紧支撑的多项式多小波, 其逼近阶为  $r$ 。

## 3 V 正交基网络

小波网络与正交基网络是基于 BP 算法的前馈神经网络的变形, 反向传播(BP)算法是最小均方(LMS)算法的推广, 它可以用来训练多层网络, BP 算法的本质是最速下降算法的近似, 与 LMS 不同的是 BP 算法使用链式法则计算各层导数<sup>[1]</sup>。下面考虑用  $r-1$  次 V 系统的基函数作为 BP 网络的隐层神经元, 构造基于 BP 算法的 V 正交基网络。

### 3.1 网络模型

假设  $f \in L^2([0, 1])$ , 由定理 1 有  $f = \sum_{n=0}^{+\infty} a_n v_n$ , 由 Riemann-Lebesgue 引理可知:  $\lim_{n \rightarrow \infty} a_n = 0$ , 因此,  $f$  可用 Fourier-V 级数的有限项和来近似表示, 即

$$f \approx \sum_{k=0}^{n-1} a_k v_k \quad (2)$$

式中,  $a_k = \int_0^1 f(x) v_k(x) dx$ 。另一方面, 系数  $a_k$  也可以通过

BP 算法学习获得, 这样可以得到对应的网络模型, 即

$$f \approx \sum_{k=0}^{n-1} w_k v_k \quad (3)$$

式(3)与式(2)唯一不同的是式(3)中权系数是通过学习得到, 而式(2)中的系数是 Fourier-V 系数。由此可以得到一个三层结构的 BP 神经网络, 如图 1 所示, 其中输入层与输出层都使用线性函数作为传递函数, 而隐层神经元则使用 V 系统的基函数作为激励函数, 由于  $v_0 = 1$ , 因此  $v_0$  相当于网络输出层的偏置。

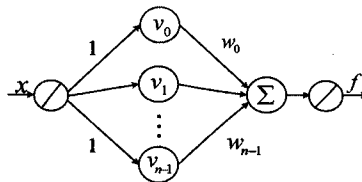


图 1 V 正交基网络模型

### 3.2 学习算法

假设训练样本集为  $\{(x_1, t_1), \dots, (x_R, t_R)\}$ ,  $t_k$  为目标值, 每次样本学习的输出值为  $f_k = \sum_{n=0}^{N-1} w_n v_n(x_k)$ , 记每次学习的网络输出与目标值之间的误差为  $e_k = t_k - f_k$ , 网络训练的目的在于通过学习调整权值, 使网络输出值与目标值的均方误差的数学期望达到最小值, 即

$$F(x) = \frac{1}{2} E[(t - f)^2] \quad (4)$$

达到最小值,  $x$  是输入(随机变量),  $t$  是目标输出,  $f$  是网络输出,  $E[\cdot]$  表示数学期望。Widrow-Hoff 学习规则的基本思想是用第  $k$  次的均方误差代替均方误差的期望<sup>[1]</sup>, 即网络训练的性能函数可以简化为  $F = \frac{1}{2} e_k^2$ 。假设学习速率为  $\eta$ , 由 LMS 算法, 可以得到权值的更新规则

$$w_n(j+1) = w_n(j) + \eta e_k v_n(x_k) \quad (5)$$

用文献[1]类似的分析方法, 由式(4)与式(5)可以得到 V 正交基神经网络收敛时的学习率应满足的条件, 即  $0 < \eta < 2/\lambda_{\max}$ , 其中  $\lambda_{\max}$  为  $R$  的最大特征值,  $R = E[vv^T]$ ,  $v = (v_0(x) \ v_1(x) \ \dots \ v_{n-1}(x))^T$ 。

### 3.3 权值直接计算

事实上, 用给定的样本训练网络的目的是寻找适当的权值, 使得式(4)的值达到最小, 由线性回归原理, 可以直接导出权值计算公式

$$w = (V^T V)^{-1} V^T t \quad (6)$$

其中,

$$w = (w_0 \ w_1 \ \dots \ w_{n-1})^T, t = (t_1 \ t_2 \ \dots \ t_R)^T$$

$$V = \begin{bmatrix} v_0(x_1) & v_1(x_1) & \dots & v_{N-1}(x_1) \\ v_0(x_2) & v_1(x_2) & \dots & v_{N-1}(x_2) \\ \dots & \dots & \dots & \dots \\ v_0(x_R) & v_1(x_R) & \dots & v_{N-1}(x_R) \end{bmatrix}$$

这个结果与文献[5-7]的伪逆规则的结果是相同的。虽然可以直接计算权值, 但是计算一个矩阵的逆矩阵并不是一件很容易的事情, 因此, 在某些实际应用中, 宁可使用 BP 网络计算权值, 也不愿意采用式(6)直接计算权值。

## 4 函数逼近的数值实验

函数逼近是神经网络的一个主要应用, 也是测试神经网络

网络收敛性能的主要手段,下面考虑用 V 正交基网络逼近一元连续函数与间断函数,为了验证 V 正交基网络的性能,分别与传统的 BP 网络、小波网络<sup>[2]</sup>、Legendre 网络<sup>[5]</sup>进行对比。

#### 4.1 实验条件

为了便于对比,在网络训练过程中,设置目标误差为 0,其目的在于:在相同训练次数下,比较 4 种网络的逼近性能,绘制重构图形与原图形的对比曲线。采用相同的网络结构(如图 1 所示),其中隐层神经元数目为 16,学习速率均为 0.001,并使用相同测试样本与仿真样本。

在传统的 BP 网络中,隐层神经元为 logsig 函数,即  $u(x) = \frac{1}{1+e^{-x}}$ ,输入层与输出层的传递函数为线性函数(purelin),网络训练算法为尺度共轭梯度反向传播算法(Scaled Conjugate Gradient Backpropagation),在 Matlab7.0 中对应的训练函数为 trainscg。小波网络的隐层神经元为高斯函数的导数,即  $\psi(x) = -xe^{-\frac{1}{2}x^2}$ ,学习算法为随机梯度算法<sup>[2]</sup>。Legendre 网络的神经元为  $[0, 1]$  区间上的 Legendre 多项式,学习算法为 LMS 算法<sup>[5]</sup>。

V 正交基网络的隐层神经元为  $[0, 1]$  区间上 3 次 V 系统(V-3)的基函数,由文献[9]与式(1)可知,3 次 V 系统的基函数可以统一表示为

$$v_n(x) = \begin{cases} v_n^{(0)}(x), & n=0,1,2,3 \\ v_n^{(1)}(x), & n=4,5,6,7 \\ \sqrt{2^j} v_p(2^j x - k), & n \geq 8 \end{cases}$$

式中,  $j = \lceil \log_2 \frac{n}{4} \rceil$ ,  $k = \lfloor \frac{n}{4} \rfloor - 2^j$ ,  $p = n - 4(2^{j-1} + k)$ ,  $v_0^{(0)} = 1$ ,  $v_1^{(0)} = \sqrt{3}(1-2x)$ ,  $v_2^{(0)} = \sqrt{5}(6x^2-6x+1)$ ,  $v_3^{(0)} = \sqrt{7}(-20x^3+30x^2-12x+1)$ ,  $v_4^{(1)}|_{0 \leq x < \frac{1}{2}} = \sqrt{7}(-64x^3+66x^2-18x+1)$ ,  $v_5^{(1)}|_{0 \leq x < \frac{1}{2}} = \sqrt{5}(-140x^3+114x^2-24x+1)$ ,  $v_6^{(1)}|_{0 \leq x < \frac{1}{2}} = \sqrt{3}(-224x^3+156x^2-28x+1)$ ,  $v_7^{(1)}|_{0 \leq x < \frac{1}{2}} = -280x^3+180x^2-30x+1$ 。

而  $v_k^{(1)}$ ,  $k=0,1,2,3$  的右半部分可由函数生成元的对称性得到。

#### 4.2 实验结果

考虑  $[0, 1]$  区间上的 3 个函数

$$f(x) = 1 + \sin(4\pi x)$$

$$g(x) = \begin{cases} e^{2x-1} \sin(4\pi x), & 0 \leq x < \frac{1}{2} \\ 8x^2 - 16x + 7, & \frac{1}{2} \leq x \leq 1 \end{cases}$$

$$h(x) = \begin{cases} 32x^2 - 12x + 1, & 0 \leq x < \frac{1}{4} \\ 2\sin(10x-2) + 1, & \frac{1}{4} \leq x < \frac{1}{2} \\ 3e^{3x-2} + 2, & \frac{1}{2} \leq x < \frac{3}{4} \\ 40x^2 - 75x + 34, & \frac{3}{4} \leq x \leq 1 \end{cases}$$

图 2、图 3 与图 4 是  $f(x)$ 、 $g(x)$ 、 $h(x)$  分别训练 20 次后得到的仿真结果,由此可以看出,在函数逼近方面,V 正交基网络的收敛速度要明显地优于传统的 BP 网络、小波网络与 Legendre 网络,特别地,对于小波网络与 Legendre 正交基网络来说,一般无法逼近分段点在 2 进制有理数点处的间断函

数。表 1 是 BP 神经网络、小波网络、Legendre 正交基网络与 V 正交基网络训练 20 次后得到的均方误差,即  $E_{rr} = \frac{1}{N} \sum_{k=1}^N [t_k - y_k]^2$ ,其中,  $t_k$  为目标值,  $y_k$  为仿真值。

表 1 4 种类型的神经网络训练 20 次后的仿真均方误差

| $E_{rr}$ | BP     | Wavelet | Legendre | V-3                     |
|----------|--------|---------|----------|-------------------------|
| $f(x)$   | 0.2520 | 0.0341  | 0.4063   | $6.9764 \times 10^{-4}$ |
| $g(x)$   | 0.0906 | 0.0363  | 0.2217   | 0.0013                  |
| $h(x)$   | 1.1927 | 0.7883  | 3.5918   | 0.0687                  |

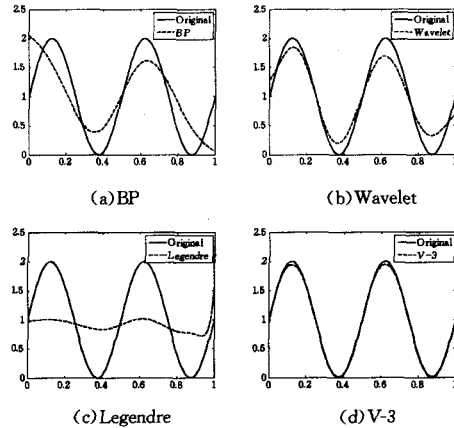


图 2 BP、Wavelet、Legendre、V-3 网络逼近连续函数  $f(x)$ , 训练次数为 20

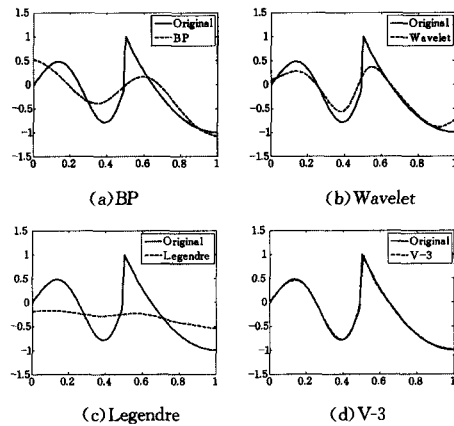


图 3 BP、Wavelet、Legendre、V-3 网络逼近间断函数  $g(x)$ , 训练次数为 20

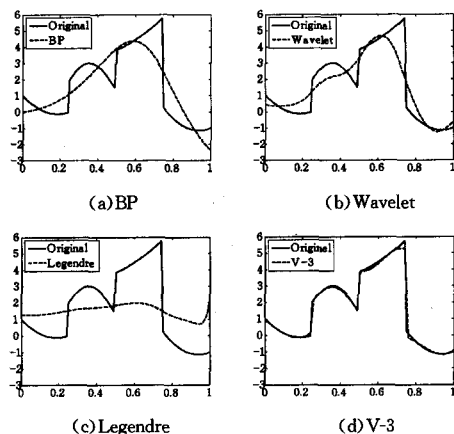


图 4 BP、Wavelet、Legendre、V-3 网络逼近间断函数  $h(x)$ , 训练次数为 20

结束语 本文研究了 V 系统在 BP 神经网络中的应用,

即用 V 系统的基函数作为前馈神经网络中的隐层神经元,构造基于 BP 算法的 V 正交神经网络,在一元函数逼近方面其收敛性能优于传统的 BP 神经网络、小波网络与 Legendre 正交神经网络。

## 参考文献

[1] Hagan M T, Demuth H B, Beale M H. Neural Network Design [M]. 戴葵,译.北京:机械工业出版社,2002

[2] Zhang Qing-hua, Benveniste A. Wavelet Networks [J]. IEEE T Neural Networ, 1992, 3(6): 889-898

[3] Zhang Jun, Walter G G, Miao Yu-bo, et al. Wavelet neural networks for function learning [J]. IEEE T Signal Proces, 1995, 43(6): 1485-1491

[4] 何新贵, 许少华. 过程神经网络[M]. 北京: 科学出版社, 2007

[5] 张雨浓, 钟童科, 李巍, 等. Legendre 正交基前向神经网络及其

其权值直接确定法[J]. 暨南大学学报: 自然科学版, 2008, 29(3): 249-253

[6] 张雨浓, 陈裕隆, 姜孝华, 等. 一种权值直接确定及结构自适应的 Chebyshev 基函数神经网络[J]. 计算机科学, 2009, 36(6): 210-213

[7] 肖秀春, 张雨浓, 姜孝华, 等. 第二类 Chebyshev 前向神经网络权值直接确定及结构自适应确定[J]. 大连海事大学学报, 2009, 35(1): 80-84

[8] Song Rui-Xia, Ma Hui, Wang Tian-Jun, et al. Complete orthogonal V-system and its applications[J]. Commun Pur Appl Anal, 2007, 6(3): 853-871

[9] Feng Yu-yu, Qi Dong-xu. A sequence of piecewise orthogonal polynomials[J]. SIAM J Math Anal, 1984, 15(4): 834-844

[10] 熊刚强. 一类分段多项式正交函数系的构造与应用[D]. 广州: 中山大学, 2010

(上接第 204 页)

在以上条件下得到仿真区域聚合及构建维边界有序标表示的时间变化趋势,如图 3 所示。在  $N=10, \lambda=10$  时,仿真区域聚合及构建边界有序表时效率最高。

由于本算法为发布成员本地区域匹配,故随机生成 1000 个联邦发布成员,分别同上面生成的联邦订购成员进行区域匹配,取平均时间值作为区域匹配时间,得到平均匹配时间,如图 4 所示。分段数为 0,  $\lambda$  为无穷时,不进行区域聚合的平均匹配时间较长,分段数少。 $\lambda$  值设置较大时,平均匹配效率较高。由于分段数少,  $\lambda$  值大时,区域聚合度较大,会直接导致区域数量减少,进而匹配次数也随之减少,故效率较高。但由于对原来的区域有所扩大,因此会有部分虚假链接,但这不会出现丢失。

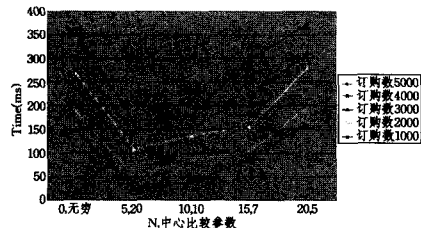


图 4 平均匹配时间

在以上构建好的订购区域中,设定  $N=10, \lambda=10$ ,随机分别选择 5%、5%~10%、10%~20%、20%~50%、50%~100% 的区域进行更新,得到平均匹配时间的仿真结果如图 5 所示。

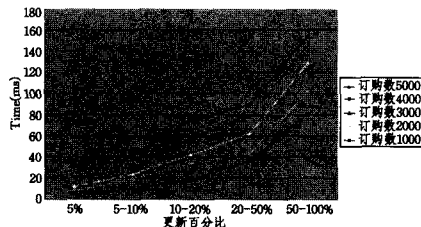


图 5 更新时间

最后比较经典排序区域匹配算法和本文提出的改进排序匹配算法的整体效率。在仿真过程中,设置改进排序匹配算法中每维分段数  $N=10$ ,中心比较参数  $\lambda=10$ ,随机在二维路径空间中 1%~100% 区域内,随机生成 5000 个订购区域,生成一个本地发布区域,随机分别选择 5%、5%~10%、10%~

20%、20%~50%、50%~100% 的订购区域进行更新。对于每次情况模拟 100 次,取其平均值为仿真结果,并进行对比。通过图 6 可以看出,改进后的算法在整体效率上相较于经典算法有了较大的提升。

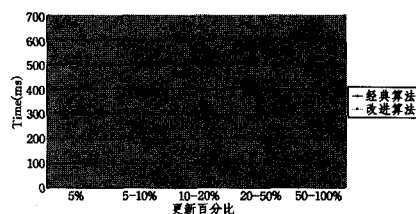


图 6 经典和改进算法的整体效率时间比较

**结束语** 本文在直接区域匹配算法的基础上,引入区域聚合的思想,同时将历史信息排序加入到区域匹配算法中,通过有效利用历史信息简化区域相交索引表和相交信息表的方式来提高区域匹配的运行效率。在仿真实验中考察了算法的维边界有序表构建时间、平均匹配时间和维边界有序表更新时间,并以这 3 个时间为考察指标,实验结果表明所提算法在时间效率上具有优势。

## 参考文献

[1] Racz C, Tan G, Yu J. A Sort-based DDM Matching Algorithm for HLA [J]. ACM Transactions on Modeling and Computer Simulation, 2005, 15(1): 14-38

[2] Morse K L, Petty M D. Data Distribution Management Migration from DoD 1. 3 to IEEE 1516 [C] // Proceedings of the 5th IEEE International Workshop on Distributed Simulation and Real Time Applications. Aug. 2001: 58-65

[3] Van Hook D, Calvin J. Data Distribution Management in RTI [C] // Spring Simulation Interoperability Workshop, No. 98S-SIW-206. 1998: 78-87

[4] 王磊, 张慧慧, 李开生, 等. 基于动态 R-树结构的 DDM 区域匹配算法[J]. 计算机工程, 2008, 34(3): 56-58

[5] 刘胤田, 唐常杰, 曾涛, 等. 基于空间填充曲线的数据分发区域匹配[J]. 系统仿真学报, 2007, 19(4): 63-71

[6] 张家祥. HLA 仿真技术应用教程[M]. 北京: 国防工业出版社, 2007: 86-97

[7] Racz C, Tan G, Yu J. A Sort-based DMM Matching Algorithm for HLA [J]. ACM Transactions on Modeling and Computer Simulation, 2005, 15(1): 96-110