

刊,2013,12(5):54-56.

数值优化改进的 BP 网络的模式分类对比^①

丁 硕, 常晓恒, 巫庆辉

(渤海大学 工学院, 锦州 121013)

摘 要: 研究了共轭梯度算法、拟牛顿算法、LM 算法三类常用的数值优化改进算法, 基于这三类数值优化算法分别对 BP 神经网络进行改进, 并构建了相应的 BP 神经网络分类模型, 将构建的分类模型应用于二维向量模式的分类, 并进行了泛化能力测试, 将不同 BP 网络分类模型的分类结果进行对比. 仿真结果表明, 对于中小规模的网络而言, LM 数值优化算法改进的 BP 网络的分类结果最为精确, 收敛速度最快, 分类性能最优; 共轭梯度数值优化算法改进的 BP 网络的分类结果误差最大, 收敛速度最慢, 分类性能最差; 拟牛顿数值优化算法改进的 BP 网络的分类结果误差值、收敛速度及分类性能介于上述两种算法之间.

关键词: 数值优化; BP 神经网络; 模式分类; 泛化能力; 对比

Comparative of Pattern Classification of BP Neural Networks Improved by Numerical Optimization Approach

DING Shuo, CHANG Xiao-Heng, WU Qing-Hui

(College of Engineering, Bohai University, Jinzhou 121013, China)

Abstract: Three common numerical optimization algorithms are first studied, including conjugate gradient algorithm, quasi-newton algorithm and LM algorithm. The three kinds of algorithms are used to improve BP neural network respectively and the corresponding classification models based on BP neural network are established. Then the models are used in pattern classification of two-dimensional vectors, and their generalization abilities are also tested. The classification results of different classification models based on BP network are compared with each other. Simulation results show that for small or medium scale networks, BP neural network improved by LM algorithm has the most precise classification result, the fastest convergence speed and the best classification ability. The one improved by conjugate gradient algorithm has the biggest error, slowest convergence speed and worst classification ability. While the classification precision, convergence speed and classification ability of quasi-newton algorithm lie between the above two algorithms.

Key words: numerical optimization; BP neural networks; pattern classification; generalization ability; comparison

射能力、大规模并行协同处理能力、鲁棒性和容错性, 这为构造神经网络模式分类系统提供了有力手段. 在
神经网络具有较强的自学习能力、非线性映 众多神经网络模型中, 应用得最为广泛的模型为 BP

神经网络. 标准的 BP 模型使用梯度下降算法进行训练, 因此 BP 神经网络的结构必须是事先已知的, 而且该学习算法收敛速度很慢, 并有可能收敛于局部最小

① 基金项目: 国家自然科学基金(61104071)

收稿时间: 2013-09-26; 收到修改稿时间: 2013-10-16

点. 这些不利因素限制了 BP 模型在模式分类中的应用. 所以, 很多学者提出了许多改进 BP 算法. BP 网络的算法, 其实质是对非线性函数进行优化, 而以往对于此类问题的求解, 经常用到数值优化的方法, 数值优化算法需要同时考虑非线性函数的一阶导数和二阶导数信息^[1-3], 这使得数值优化的方法在求解非线性函数时具有收敛速度较快的优点, 所以一部分学者利用

数值优化的方法对 BP 算法进行改进. 尽管这些研究成果都在不同程度上改善了 BP 网络性能, 但是究竟何种数值优化算法改进的 BP 网络更适合于模式分类, 网络收敛速度更快且分类结果更为准确, 涉及到这些问题的文献尚不是很多. 本文详细分析了三类基于数值优化改进的 BP 算法, 在 MATLAB7.0 环境下, 基于共轭梯度算法、拟牛顿算法、LM 算法三种常用的数值优化改进算法分别构建了相应的 BP 神经网络, 并将其应用于二维向量模式的分类, 通过实例仿真, 发现各类数值改进方法的特性, 比较各类数值优化算法应用于 BP 神经网络后的分类效果, 得出在不同条件下, 每种数值优化算法的分类的可行性及其各自的优越性.

1 数值优化改进的BP神经网络算法原理

1.1 共轭梯度优化算法原理

共轭梯度优化算法的实质是沿着梯度最陡的方向下降来对权值进行修正, 误差函数沿着梯度最陡的方向下降最快, 收敛速度比沿着梯度最陡的方向的收敛速度更快. 典型共轭梯度优化算法主要有 4 种不同优化算法, 分别是: Fletcher-Reeves、Polak-Ribiere、Powell-Beale 和 Scaled Conjugate Gradient(SCG)优化算

法. 共轭梯度优化算法的首轮迭代过程如式(1)所示:

$$p(0) = -g(0) \quad (1)$$

最佳线性搜索方向如式(2)所示:

$$x(k+1) = x(k) + \alpha p(k) \quad (2)$$

$$p(k) = -g(k) + \beta(k)p(k-1) \quad (3)$$

式(3)中, $p(k)$ 为经过 $k+1$ 轮迭代后的搜索方向, 它由前一轮迭代后的梯度、搜索方向共同决定, $\beta(k)$ 的算法视其具体属于哪一种共轭梯度优化算法而定. 在 MATLAB 的 Toolbox 中, Fletcher-Reeves 优化算法对应的训练函数为 `traincgf` ($\beta(k)$ 的算法如式(4)所示)和 Polak-Ribiere 优化算法的对应的训练函数为 `traincgp` ($\beta(k)$ 的算法如式(5)所示)^[4].

$$\beta(k) = (g^T(k)g(k)) / (g^T(k-1)g(k-1)) \quad (4)$$

$$\beta(k) = (\Delta g^T(k-1)g(k)) / (g^T(k-1)g(k-1)) \quad (5)$$

在 Powell-Beale 优化算法的迭代过程如式(6)和式(7)所示:

$$p(k) = -g(k) \quad (6)$$

$$|g^T(k-1)g(k)| \geq 0.2 \|g(k)\|^2 \quad (7)$$

SCG 优化算法在 4 种共轭梯度优化算法中迭代过程最为简单, 计算量最小. 在 MATLAB 的 Toolbox 中, Powell-Beale 优化算法对应的训练函数为 `traincgb`; SCG 优化算法对应的训练函数为 `trainscg`.

1.2 拟牛顿优化算法原理

拟牛顿法是一种改进 BP 算法的常见数值优化方法, 它的理论基础是牛顿法. 拟牛顿算法的权值更新如式(8)所示:

$$w(k+1) = w(k) + \eta_n p(k) \quad (8)$$

$$p(k) = -S(k)g(k) \quad (9)$$

式(9)中, 方向向量 $p(k)$ 用梯度向量 $g(k)$ 来定义, 矩阵 $S(k)$ 是每次迭代中调整的正定矩阵, 这样做目的在于方向向量 $d(k)$ 逼近牛顿法的方向 $\{H(k)\}^{-1} \times \nabla f(x(k))$. 令 $q(k) = g(k+1) - g(k)$, $\Delta w(k) = w(k+1) - w(k)$, 且通过逼近式(10)可以成立.

$$q(k) \approx (\partial g(k) / \partial w) \times \Delta w(k) \quad (10)$$

用 $S(k)$ 去逼近 $H(k)$, 则式(11)可以成立.

$$S(k) \approx H(k) = (\partial g(k) / \partial w) \approx q(k) \times \Delta w^{-1}(k) \quad (11)$$

当误差函数是二次函数时, 式(11)是精确的, 拟牛顿算法实现了对 Hessian 矩阵 $H(k)$ 的逼近. $H(k)$ 的近似

矩阵 $S(k)$ 根据式(11)使用递归公式可得式(12):

$$S(k+1) = S(k) + (\Delta w(k) \Delta w^T(k)) / (q^T(k) q(k)) - (S(k) q(k) q^T(k) S(k)) / (q^T(k) S(k) q(k)) + \varepsilon(k) [q^T(k) S(k) q(k)] [v(k) v^T(k)] \quad (12)$$

式(12)中, $v(k) \approx (\Delta w(k)) / (\Delta w^T(k) \Delta w(k)) - (S(k) q(k)) / q^T(k) S(k) q(k) \quad 0 \leq \varepsilon \leq 1$

当式(12)中 $\varepsilon(k)$ 取值不同时, 逼近 Hessian 矩阵的方法不同^[5]. 比较典型的是 BFGS(Broyden, Fletcher, Goidfard, Shanno)优化算法和一步割线优化算法. 在 MATLAB 的 Toolbox 中, 采用 BFGS 优化算法和一步割线优化算法的训练函数分别为 `trainbfg` 和 `trainoss`.

1.3 LM(Levenberg-Marguardt)优化算法原理

LM 算法与拟牛顿算法一样是为了在近似二阶训练速率进行修正时避免计算 Hessian 矩阵而设计的. 若误差函数用平方和表示时, 可以用式(13)来表示 Hessian 矩阵:

$$H(k) = J^T(k) J(k) \quad (13)$$

梯度表达式如式(14)所示.

$$g = J^T(k) e(k) \quad (14)$$

式(13)中, $J(k)$ 是包含网络误差对权值和阈值的一阶导数的雅克比矩阵, 式(14)中, e 为误差向量. LM 算法可按照式(15)进行修正:

$$x(k+1) = x(k) - [J^T J + \mu I]^{-1} J^T e(k) \quad (15)$$

式(15)中, I 为单位矩阵, 比例系数 μ 是一个大于 0 的很小的参数, 当 μ 接近零时, 式(15)变为牛顿法; 当 μ 很大时, 式(15)变为梯度下降法. 因为牛顿法在对最小误差进行逼近时, 收敛速度快, 精度高, 所以应使式(15)最终接近于牛顿法, 使 μ 减小; 仅在进行尝试性迭代后的误差性能增加的情况下, 才使增加^[6]. 在 MATLAB 的 Toolbox 中, LM 算法对应的训练函数为 `trainlm`.

2 数值优化改进的BP网络分类模型的建立

建立数值优化改进的 BP 神经网络, 主要包含网络层数、隐层神经元个数、初始权值和学习率四个要素.

2.1 网络层数的确定

模式分类问题实质是模式特征空间的划分问题, 也可以说是一种映射问题. 而 BP 网络的输入输出关

系可以看成是一种高度非线性的映射关系, 其信息处理能力基于简单的非线性函数的多次复合. 由 BP 网络构造定理, 对于任何在闭区间内的一个连续函数都可以用一个隐含层的 BP 网络来逼近, 即一个三层的 BP 网络可以完成任意的从 n 维到 m 维的映射^[7,8]. BP 网络样本模式分类就是基于此定理. 因此, 本文在进行模式分类实验时, BP 神经网络采用单隐层结构.

2.2 隐含层节点数的确定

隐含层的节点数目与输入输出单元的多少有关. 隐含层的节点数目过多或过少都会对网络性能产生影响. 节点数目过多会使网络规模过于庞大, 造成训练艰难; 节点数目过少会造成网络容错性差. 隐含层的神经元数目与输入、输出层神经元数目相关. 本文采用动态法来确定隐含层神经元数, 即一开始选用较少的隐层神经元, 如果学习一定次数后效果不好, 再增加隐层神经元, 一直达到比较合理的隐层神经元数为止, 经过反复多次试验隐含层神经元数最终确定为 13, 可以达到逼近要求.

2.3 初始权值的选取

初始权值对于非线性的 BP 神经网络的收敛程度和训练时间影响很大, 初始权值在输入累加时使每个神经元的状态值接近于零. 若初始权值过大, 会造成输入值在加权后落入传递函数的饱和区, 进而使得训练停止. 所以, 本文在建立几种数值优化 BP 神经网络时, 输入值在加权处理后尽可能接近零, 这样可以保证初始权值的调整发生在 S 型传递函数的斜率最陡处.

2.4 学习率

学习率过大, 系统就会不稳定; 相反, 网络的训练时间就会延长, 收敛速度缓慢, 最终陷入局部最小值, 学习率一般选取(0.01, 0.8)之间. 本文在建立几种数值优化 BP 神经网络时, 为了兼顾系统稳定性和具有较快的收敛速度, 学习率选取为 0.1.

2.5 样本的预处理

由于文中的训练样本集各元素的取值范围不完全一致, 这不仅加大了逼近函数的波动性, 使所构建的网络的训练速度下降, 容易造成网络训练失败. 因此, 先对训练样本集各元素进行预处理是十分必要的, 它在一定程度上可以提高网络的训练速度, 提高训练的成功率. 本文采用 `minmax` 函数将训练样本集中各元素归一化到(0-1)范围内.

3 仿真实验

利用数值优化算法改进的 BP 神经网络对图 1 所示的 15 个输入二维向量的模式进行分类. 网络输入层节点数被确定为 15. 在对输入二维向量的模式进行分类实验时, 网络的输出分别用(0 0)表示第一类模式、(0 1)表示第二类模式、(1 0)表示第三类模式、(1 1)表示第四类模式. 所以网络输出层节点数被确定为 2.

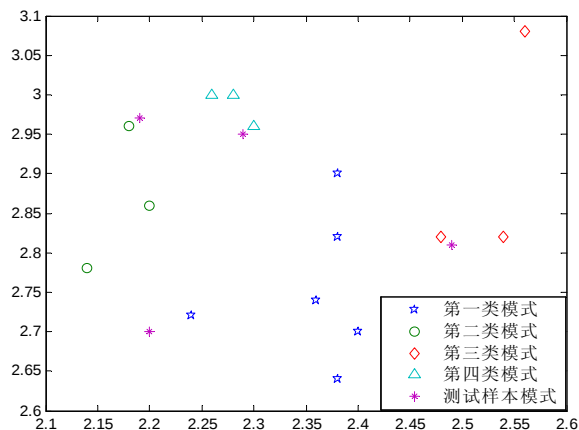


图 1 输入二维向量与测试样本向量

3.1 仿真步骤

为了便于对共轭梯度算法、拟牛顿算法、LM 算法三种常用的数值优化算法改进的 BP 神经网络的分类性能进行研究, 本文在 MATLAB7.0 环境下, 分别编程构建了相应算法的 BP 网络模型, 并对输入的二维向量模式进行分类实验, 并进行对比研究. 具体仿真过程如下^[7,8]:

① 以待分类的二维向量作为神经网络的训练样本集, 以已知二维向量的模式聚类结果作为目标样本集;

② 对共轭梯度算法(Fletcher-Reeves 算法、Polak-Ribiere 算法、Powell-Beale 算法以及 SCG 算法)、拟牛顿法(BFGS 算法和一步割线算法)和 LM 算法进行编程建立七种 BP 网络, 并用上一步骤所形成的训练样本集反复训练, 直到满足要求;如果所建立的网络经充分训练后仍无法达到要求, 则需调整网络结构;

③ 利用所建立的 BP 网络对训练样本集进行仿真;

④ 分析三大类共七种数值优化改进 BP 网络的仿真结果, 以确定在解决相同分类问题时, 每一类数值优化算法当中分类性能较为优秀的一种算法.

⑤ 检验所建立的 BP 网络的泛化能力. 利用上一

步骤确定的三种算法所建立的 BP 网络对测试样本进行仿真(不在训练样本空间内的样本), 比较其仿真误差;

⑥ 在训练样本数相同且精度要求相等的条件下, 对于三种网络的分类结果进行比较, 分析不同类型的数值优化改进 BP 网络的分类性能.

3.2 改进的 BP 网络分类模型对训练样本的分类结果

在目标精度设置为 0.0001、训练样本数目相等的条件下, 7 种数值优化改进 BP 算法仿真结果对比如表 1 所示.

可以看出, 在 4 种共轭梯度优化算法中, Fletcher - Powell 算法需要 146 步才能达到目标精度, 训练时间最长, 均方误差最大, SCG 算法和 Powell - Beale 算法所需训练时间稍短, 均方误差略微减小, Polak - Ribiere 算法需要 26 步就可以达到目标精度, 训练时间最短, 均方误差最小; 在 2 种拟牛顿优化算法中, 一步割线算法需要 120 步才能达到目标精度, 训练时间最长, 均方误差最大, 而 BFGS 算法只需 23 步即可达到目标精度, 训练时间最短, 均方误差最小. 总的来说,

表 1 数值优化改进 BP 算法仿真结果对比

训练函数	BP 算法名称	收敛步数	均方误差
traincgf	Fletcher - Powell 算法	146	9.95272e-005
	SCG 算法	76	9.41881e-005
traincgb	Powell - Beale 算法	53	9.16615e-005
	Polak - Ribiere 算法	26	9.00932e-005
trainoss	一步割线算法	120	9.85183e-005
trainbfg	BFGS 算法	23	5.77583e-005
trainlm	LM 算法	9	2.60203e-005

相对于训练样本集而言, 7 种数值优化改进 BP 算法都能在规定的训练步数范围内完成对已知二维向量的模式聚类任务, 但相比之下, 在 4 种共轭梯度优化算法中, Polak - Ribiere 算法分类性能最优; 在 2 种拟牛顿优化算法中, BFGS 算法分类性能最优; Polak - Ribiere 算法和 BFGS 算法所需训练时间大致相同, 但 BFGS 算法均方误差明显比 Polak - Ribiere 算法小. LM 算法只需 23 步即可达到目标精度, 网络训练时间远远少于

共轭梯度优化算法和拟牛顿优化算法, 均方误差也远小于共轭梯度优化算法和拟牛顿优化算法, 是所有 7 种数值优化改进 BP 算法中所需训练时间最短、均方误差最小的一种优化算法。

3.3 改进的 BP 网络分类模型的泛化能力测试

为了进一步检验基于共轭梯度优化算法、拟牛顿优化算法和 LM 优化算法改进的 BP 网络分类器的泛化能力, 文中分别利用三类优化算法中分类性能最为优越的 Polak - Ribiere 算法、BFGS 算法和 LM 算法所改进的 BP 网络(经过训练后的网络)对测试样本(不在训练样本空间内的样本)进行分类, 设置 4 个特定的分属于不同类别的向量, 如图 1 所示, (2.20,2.70) 属于第 1 类, (2.19,2.97) 属于第 2 类, (2.49,2.81) 属于第 3 类, (2.29,2.95) 属于第 4 类, 利用仿真函数得到 BP 神经网络的输出。

Polak - Ribiere 算法、BFGS 算法和 LM 算法改进的 BP 网络对测试样本的分类测试误差变化曲线如图 2、图 3、图 4 所示。从图 2-图 4 可以看出, 在上述三种数值优化算法中, Polak - Ribiere 算法需要 35 步才能达到目标精度, 收敛时间最长, 均方误差最大, BFGS 算法所需训练时间稍短, 需要 24 达到目标精度, 均方误差有所减小, LM 算法只需 9 步即可达到目标精度, 收敛时间最短, 均方误差也最小。三种数值优化改进的 BP 网络对测试样本的分类结果对比如表 2 所示, 从表 2 可以看出, LM 算法最大绝对误差为-0.0042, BFGS 算法最大绝对误差为-0.1004, Polak - Ribiere 算法最大绝对误差为 0.1091。

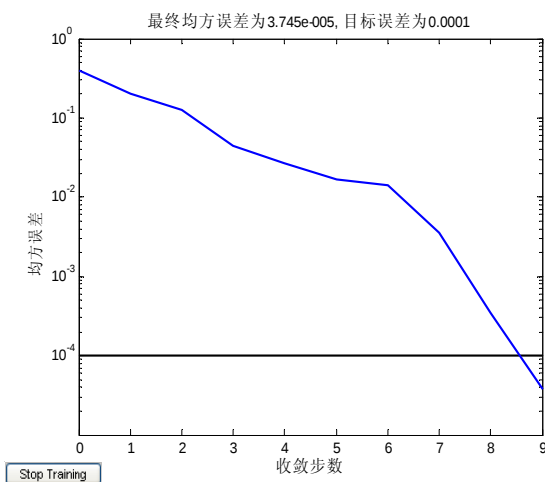


图 2 改进的 BP 网络分类测试误差变化曲线(trainlm)

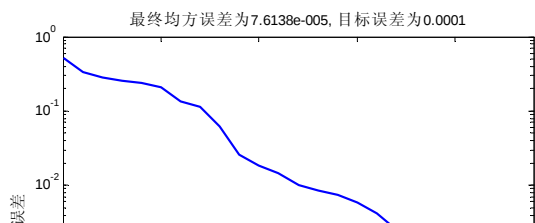


图 3 改进的 BP 网络分类测试误差变化曲线(trainbfg)

总的来说, 相对于测试样本集而言, 三种数值优化算法改进 BP 网络都能在规定的训练步数范围内完成对测试样本的模式分类任务, 但相比之下, Polak - Ribiere 算法分类结果误差值最大, 收敛时间最长; BFGS 算法分类结果误差值较 Polak - Ribiere 算法小, 收敛时间较 Polak - Ribiere 算法短; LM 算法对于测试样本的分类结果的绝对误差明显小于 Polak - Ribiere

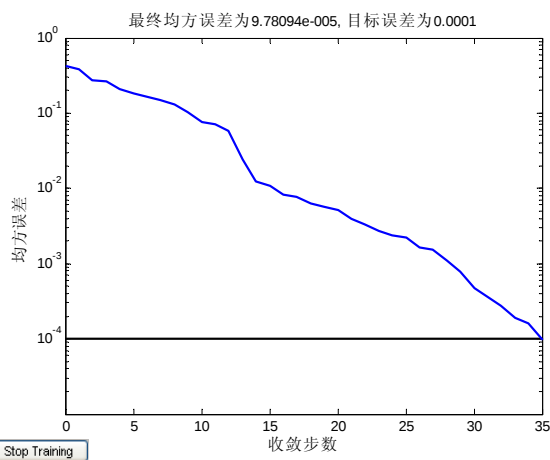


图 4 改进的 BP 网络分类测试误差变化曲线(traincgp)

算法和 BFGS 算法, 只是在个别测试样本点的绝对误差略有增大但仍远小于 Polak - Ribiere 算法和 BFGS 算法, 几乎做到了完全逼近且收敛时间最短。因此可以说, 对于中小规模的网络而言, LM 数值优化算法改进的 BP 网络的分类结果最为精确, 收敛速度最快, 分类性能最优; 共轭梯度数值优化算法改进的 BP 网络的分类结果误差最大, 收敛速度最慢, 分类性能最差; 拟牛顿数值优化算法改进的 BP 网络的分类结果误差值、收敛速度及分类性能介于上述两种算法之间。

表 2 三种数值优化改进的 BP 网络
对测试样本的分类结果对比

训练 函数	目标 输出	实际输出		绝对误差	
trainlm	0 0	0.0000	0.0042	0.0000	0.0042
	0 1	0.0021	1.0000	0.0021	0.0000
	1 0	0.9994	0.0000	-0.0006	0.0000
	1 1	0.9991	0.9998	-0.0009	-0.0002
trainbfg	0 0	0.0000	0.0043	0.0000	0.0043
	0 1	0.0613	1.0000	0.0613	0.0000
	1 0	0.9893	0.0000	-0.0107	0.0000
	1 1	0.8996	0.9985	-0.1004	-0.0015
traincgp	0 0	0.0000	0.1091	0.0000	0.1091
	0 1	0.0688	1.0000	0.0688	0.0000
	1 0	0.9867	0.0051	-0.0133	0.0051
	1 1	0.8931	0.9964	-0.1069	-0.0036

4 结语

文中利用三类数值优化算法构建了改进的 BP 网络分类模型,并对输入的二维向量模式进行分类实验.仿真结果表明:共轭梯度算法、拟牛顿算法和 LM 算法都能在规定预设精度范围内完成分类任务.但相比之下,共轭梯度算法的分类效果相对较差.在四种典型的共轭梯度算法中,Polak-Ribiere 优化算法和其它三种共轭梯度算法相比,需要较短的收敛时间,勉强达到分类精度要求.采用 LM 算法收敛速度最快,如果就训练准确性和分类精度来讲,该算法明显优于其他几种数值优化算法,但 LM 算法算法的复杂度较大,当 BP 神经网络的结构相对简单且网络参数很少时,该算法比较容易实现,分类效果好;当 BP 神经网络的结构比较复杂时,该算法在计算过程中会产生大量的中间结果矩阵,需要较大的计算机内存空间,实现 LM 算法的难度会大大增加.所以在工程领域中进行设备的故障分类、故障诊断以及对经济、国土资源、环境保护等领域内的区域划分时应优先采用 LM 算法进行分类.尽管拟牛顿算法进行逼近时要求存储近似 Hessian 矩阵,但它仍比共轭梯度法的收敛速度要快,仅次于 LM 算法,所以,如果内存空间较小,也可以尝试使用拟牛顿算法.

参考文献

- 1 高雪鹏,丛爽.BP 网络改进算法的性能对比研究.控制与决策,2001,16(2):167-171.
- 2 周黄斌,周永华,朱丽娟.基于 MATLAB 的改进 BP 神经网络的实现与比较.计算技术与自动化,2008,27(1):28-31.
- 3 曹旭帆,叶舟,万俊,等.基于 BP 神经网络的函数逼近实验及 MATLAB 实现.实验室研究与探索,2008,27(5):34-38.
- 4 丁硕,巫庆辉.基于改进 BP 神经网络的函数逼近性能对比研究.计算机与现代化,2012,11:10-13.
- 5 刘天舒.BP 神经网络的改进研究及应用[学位论文].哈尔滨:东北林业大学,2011.
- 6 Ding S, Wu QH. A MATLAB-based study on approximation performances of improved algorithms of typical BP neural networks. Applied Mechanics and Materials, 2013, 313-314: 1353-1356.
- 7 周凌翱.改进 BP 神经网络在模式识别中的应用及研究[学位论文].南京:南京理工大学,2010.
- 8 江虹,曾立波,胡继明.优化的 BP 神经网络分类器的设计与实现.计算机工程与应用,2001,5:122-125.