

具有最优学习率的 RBF 神经网络及其应用^①

卫 敏², 余乐安^{1, 2}

(1. 北京化工大学经济管理学院, 北京 100029;

2. 中国科学院数学与系统科学研究院, 北京 100190)

摘要: 传统固定学习率的 RBF 神经网络在金融时间序列预测方面已经有比较成功的应用, 但网络学习率的选择问题却给传统 RBF 神经网络的使用带来了不便. 利用梯度下降法及优化方法推导出了 RBF 神经网络的动态最优学习率并将其应用于网络学习算法, 具有最优学习率的 RBF 神经网络能够在保证网络稳定学习的同时兼顾网络的收敛速度. 为了检验具有动态最优学习率的 RBF 神经网络的预测效果, 对沪深 300 指数波动率进行了预测实验. 实验结果表明, 具有动态最优学习率的 RBF 神经网络比传统的固定学习率的 RBF 神经网络有着更快的收敛速度, 同时也避免了人为选定学习率的不便.

关键词: RBF 神经网络; 最优学习率; 梯度下降法

中图分类号: F830.91 文献标识码: A 文章编号: 1007-9807(2012)04-0050-08

0 引 言

RBF 神经网络 (radial basis function neural networks, 径向基神经网络) 是一类流行的人工神经网络, 其优秀的函数逼近能力使其在模式识别^[1-4]、函数逼近^[5-8]、经济预测^[9]等领域有着较为广泛的应用. RBF 神经网络是一种具有三层结构的神经网络, 即: 输入层、隐含层和输出层. 通常, 一个具有 m 个输入值、 n 个隐含节点及一个输出值的 RBF 神经网络结构如图 1 所示.

在图 1 中, RBF 神经网络的输入层为 (x_1, x_2, \dots, x_m) , 隐含层为 $(\varphi_1(x), \varphi_2(x), \dots, \varphi_n(x))$, 输出层为 $f(x)$ 而 (w_1, w_2, \dots, w_n) 为隐含层与输出层之间的连接权重. 隐含层的每一个节点都使用非线性函数 $\varphi(\cdot)$ 作为径向基函数 (radial basis function, RBF), 也就是说, 从输入空间到隐含空间的变换是非线性的. 输入层将所有数据输入隐含层的每一个节点, 而隐含层节点经线性加权求和即得出 RBF 神经网络的输出 (权重为隐含层与

输出层的连接权重). 通常, 径向基函数 $\varphi(\cdot)$ 的选择有以下几种:

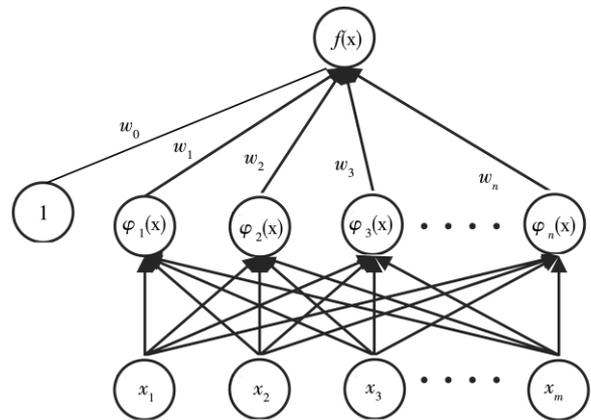


图 1 RBF 神经网络结构图

Fig. 1 A general architecture of RBF neural network

1) Gaussian 函数^[10, 11]

$$\varphi(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (1)$$

2) Reflected Sigmoidal 函数^[12]

① 收稿日期: 2011-08-26; 修订日期: 2011-09-26.

基金项目: 国家杰出青年科学基金资助项目(71025005); 国家自然科学基金重大研究计划培育项目(90924024).

作者简介: 余乐安(1976—), 男, 湖南汉寿人, 教授, 博士生导师, 通讯作者. Email: yulean@amss.ac.cn

$$\varphi(r) = \frac{1}{1 + \exp\left(\frac{r^2}{\sigma^2}\right)} \quad (2)$$

3) Multiquadrics 函数^[10]

$$\varphi(r) = (r^2 + \sigma^2)^{1/2} \quad (3)$$

这里 r 为径向基函数的半径, 而 σ^2 为径向基函数的方差.

在众多可选的径向基函数中, 人们通常选择 Gaussian 函数作为 RBF 神经网络的径向基函数. 因为 Gaussian 函数从中心到两边单调递减, 所以其回应是局部有限的, 具有更加逼真的生物特征^[13]. 与之相对比的是, Multiquadric 函数的回应是全局且无限的, 因为 Multiquadric 函数从中心到两边单调递增.

选定径向基函数之后, 通过对这些函数的结果进行线性求和, 就可以得到输出值. 这样, RBF 神经网络就完成了映射 $f: R^r \rightarrow R$, 公式如下:

$$f(x) = w_0 + \sum_{i=1}^n w_i \varphi(\|x - c_i\|) \quad (4)$$

其中 $x \in R^r$ 为输入向量, $c_i (1 \leq i \leq n)$ 为 RBF 神经网络的中心, 范数 $\|\cdot\|$ 表示欧氏距离, $w_i (1 \leq i \leq n)$ 为线性求和的权重, w_0 表示偏差 (bias). 如上文所述, 选择 Gaussian 函数作为径向基函数 $\varphi(\cdot)$. 通常, RBF 神经网络的中心 c_i 被设为不变^[14].

中心 c_i 的选择方法有三类: 1) 固定中心法. 这种方法从输入数据集中随机选择中心, 比较适用于训练数据的分布很有代表性的情况. 其缺陷是容易使网络表现不理想或使网络体积过于庞大^[10]. 2) 自组织选择法. K - 均值聚类法就是自组织中心选择法的一种, 这种方法将输入数据集进行聚类, 将聚类的中心作为网络的中心. 这种方法的优点是可以将 RBF 中心设置成相对较重要的数据点, 其缺陷是如果初始聚类的中心是准确的, 那最终结果可能会是局部最优解. 3) 有监督的中心选择法. 这种方法需要在网络学习过程中确定成本函数, 中心的确定过程就是使成本函数最小的过程^[15]. 这种方法通常要用到梯度下降法 (gradient descent). 为了避免上述几种方法所产生的问题和不便, 使用了一种更加完善的中心选择方法: 正交最小二乘法 (orthogonal least squares)^[10].

权重 w_i 需要通过反复学习 (训练) 来使神经网络的输出达到其准确性. 在本文中, 权重的计算方法为最小均方算法 (least mean square, LMS), 网络的学习算法为梯度下降法 (gradient descent). 梯度下降法是优化算法的一种^[16], 在 RBF 神经网络学习过程中, 使权重沿着与目标函数梯度相反的方向进行逐步调整, 最终求得权重的最优值, 使网络最优化. 权重的调整过程可表示为

$$w(t+1) = w(t) - \eta \frac{d}{dw} E_i(w) \quad (5)$$

其中 t 表示迭代的次数, 目标函数 $E_i(w)$ 通常为成本函数 (cost function), η 为学习率 (learning rate), 亦称搜索步长.

在 RBF 神经网络的使用中, 学习率通常被人为主观地设为一个固定值, 在整个网络学习过程中保持不变. 这样的方法存在着很多问题: 如果学习率设置过大, 网络的收敛速度可能会很快, 但也有可能会导致网络不稳定甚至无法学习; 而学习率过小又会导致网络收敛速度慢, 消耗大量的计算时间, 无法满足实际应用的时效要求. 同时, 不同的 RBF 神经网络需要选择不同的学习率, 这给 RBF 神经网络的使用带来了很大的不便. 因此, 对于传统 RBF 神经网络来说, 选择一个合适的学习率是非常困难的. 为了解决这样的问题, 将会推导一种新型的“动态最优学习率”, 这种学习率在每一步迭代都会进行优化计算, 是可变的学习率. 动态最优学习率对于每一步迭代来说都是适用的, 在保证网络稳定学习的同时兼顾网络的收敛速度, 提高网络的运行效率.

1 RBF 神经网络动态最优学习率

1.1 RBF 神经网络动态最优学习率推导

由 RBF 神经网络的结构可知, RBF 神经网络的输出为

$$f(x) = \sum_{i=1}^n w_i \varphi(\|x - c_i\|) \quad (6)$$

其中 $\varphi(\|x - c_i\|)$ 为 Gaussian 函数, w_i 为连接权重, 偏差 (bias) w_0 在这里被简化为权重的一部分. $\varphi(\|x - c_i\|)$ 可表示为

$$\varphi(\|x - c_i\|) = \exp\left(-\frac{1}{2\sigma^2}\|x - c_i\|^2\right)$$

$$i = 1, 2, \dots, n \quad (7)$$

式中 n 为网络隐含层节点的个数; c_i 为网络隐含层节点的中心; w_i 为隐含层到输出层的连接权重; σ^2 为径向基函数的方差.

设 $\varphi_{ji} = \varphi(\|x_j - c_i\|)$ ($j = 1, 2, \dots, m$; m 为输入样本总数) 可设矩阵

$$\Phi = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \dots & \varphi_{1n} \\ \varphi_{21} & \varphi_{22} & \dots & \varphi_{2n} \\ \dots & \dots & \dots & \dots \\ \varphi_{m1} & \varphi_{m2} & \dots & \varphi_{mn} \end{bmatrix}$$

则公式 (6) 可改写为:

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & \dots & \varphi_{1n} \\ \varphi_{21} & \varphi_{22} & \dots & \varphi_{2n} \\ \dots & \dots & \dots & \dots \\ \varphi_{m1} & \varphi_{m2} & \dots & \varphi_{mn} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad (8)$$

若用 $W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$ 表示权重, 用 $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$ 表示实际

输出值, 则公式 (8) 可简写为

$$\Phi W = y \quad (9)$$

用 \hat{y} 表示由网络计算而得到的输出结果, 则第 t 次学习的成本函数 (cost function) 可定义为

$$E(t) = \frac{1}{2} \sum_{j=1}^m (y_j - \hat{y}_j(t))^2$$

$$= \frac{1}{2} \sum_{j=1}^m e_j^2(t) \quad (10)$$

对成本函数 $E(t)$ 采用梯度下降法来进行处理, 由公式 (10) 可知 $E(t)$ 的梯度为

$$\begin{aligned} \frac{\partial E(t)}{\partial W(t)} &= \sum_{j=1}^m e_j(t) \frac{\partial e_j(t)}{\partial W(t)} \\ &= \sum_{j=1}^m e_j(t) \frac{\partial (y_j - \hat{y}_j(t))}{\partial W(t)} \\ &= - \sum_{j=1}^m e_j(t) \frac{\partial \hat{y}_j(t)}{\partial W(t)} \\ &= - \sum_{j=1}^m e_j(t) \frac{\partial (\Phi_j \cdot W(t))}{\partial W(t)} \\ &= - \sum_{j=1}^m e_j(t) \Phi_j^T \end{aligned} \quad (11)$$

设 $\Delta W(t) = W(t) - W(t-1)$ 为第 t 次学习时权重的改变量, 利用公式 (11) 中的梯度 $\frac{\partial E(t)}{\partial W(t)}$ $\Delta W(t)$ 可改写为

$$\begin{aligned} \Delta W(t) &= -\eta(t) \frac{\partial E(t)}{\partial W(t)} \\ &= \eta(t) \sum_{j=1}^m e_j(t-1) \Phi_j^T \end{aligned} \quad (12)$$

这里 η 表示学习率, 若用 $e(t) = \begin{bmatrix} e_1(t) \\ e_2(t) \\ \vdots \\ e_m(t) \end{bmatrix}$ 表示输

出误差, 则公式 (12) 可简写为

$$\Delta W(t) = \eta(t) \Phi^T e(t-1) \quad (13)$$

误差的增量可表示为

$$\begin{aligned} \Delta e(t) &= e(t) - e(t-1) \\ &= y(t) - \hat{y}(t) - y(t-1) + \hat{y}(t-1) \end{aligned} \quad (14)$$

设 $\Delta y(t) = y(t) - y(t-1)$ 为实际输出值的

改变量, $\Delta \hat{y}(t) = \hat{y}(t) - \hat{y}(t-1)$ 为网络输出值的改变量. 假设真实输出值改变量的绝对值远小于网络输出值改变量的绝对值, 即 $|\Delta y(t)| \ll |$

$\Delta \hat{y}(t)|$ ^[17-20], 也就是说, 实际输出值的改变量对于网络输出值的改变量来说是可以忽略的. 因为实际输出值往往受到很多条件的制约, 而网络输出值则不会受到太多的限制, 所以这样的假设具有一定的现实意义. 由此, 公式 (14) 可以近似写为

$$\Delta e(t) = e(t) - e(t-1) \approx -\Delta \hat{y}(t) \quad (15)$$

利用公式 (6) 可得误差的改变量为

$$\begin{aligned} \Delta e(t) &\approx -\Delta \hat{y}(t) \\ &= -\Phi \Delta W(t) \\ &= -\Phi \eta(t) \Phi^T e(t-1) \\ &= \eta(t) \Phi \Phi^T e(t-1) \end{aligned} \quad (16)$$

则

$$\begin{aligned} e(t) &= e(t-1) + \Delta e(t) \\ &\approx e(t-1) - \eta(t) \Phi \Phi^T e(t-1) \end{aligned} \quad (17)$$

由公式 (17) 可得第 t 次的成本函数, 即

$$\begin{aligned} E(t) &= \frac{1}{2} e^T(t) e(t) \\ &\approx \frac{1}{2} [e(t-1) - \eta(t) \Phi \Phi^T e(t-1)]^T \cdot \end{aligned}$$

$$[e(t-1) - \eta(t) \Phi \Phi^T e(t-1)] \quad (18)$$

此成本函数可以看作是学习率 η 的函数, 通过最小化 $E(t)$, 可以得到学习率 η 的最优值. 由公式 (18) 可以得出一阶条件如下

$$\begin{aligned} & \left. \frac{\partial E(t)}{\partial \eta(t)} \right|_{\eta(t)=\eta^*(t)} \\ &= -\frac{1}{2} [\Phi \Phi^T e(t-1)]^T \times \\ & \quad [e(t-1) - \eta^*(t) \Phi \Phi^T e(t-1)] - \\ & \quad \frac{1}{2} [e(t-1) - \eta^*(t) \Phi \Phi^T e(t-1)]^T \times \\ & \quad [\Phi \Phi^T e(t-1)] \\ &= 0 \end{aligned} \quad (19)$$

二阶条件为

$$\left. \frac{\partial^2 E(t)}{\partial \eta^2(t)} \right|_{\eta(t)=\eta^*(t)} = e^T(t-1) \Phi \Phi^T \Phi \Phi^T e(t-1) > 0 \quad (20)$$

因为矩阵 Φ 是正定的, 所以二阶条件即不等式 (20) 成立, 由公式 (19) 可得最优学习率为

$$\eta^*(t) = \frac{e^T(t-1) \Phi \Phi^T e(t-1)}{e^T(t-1) \Phi \Phi^T \Phi \Phi^T e(t-1)} \quad (21)$$

1.2 具有动态最优学习率的 RBF 神经网络学习算法

在推导出 RBF 神经网络动态最优学习率之后, 可以设计出具有动态最优学习率的 RBF 神经网络学习算法, 这个算法可以看作是对传统固定学习率的 RBF 神经网络学习算法的一种改进, 每一步迭代的最优学习率由公式 (21) 计算得出. 具有最优学习率的 RBF 神经网络学习算法表述如下:

步骤 1 确定网络隐含层节点数, 用正交最小二乘法^[4] 选定径向基函数中心 c 并计算方差 σ^2 , 求出矩阵 Φ . 设定学习精度 $rmse^*$, 设定网络学习的最大迭代次数 $\max t$;

步骤 2 设置隐含层至输出层的初始连接权重 W_0 , 计算网络输出值 $\hat{y} = \Phi W$, 并计算实际输出值与网络输出值的均方根误差 $rmse$;

步骤 3 判断 $rmse < rmse^*$ 是否成立或最大迭代次数 $\max t$ 是否达到, 如果两者之一成立, 进入步骤 5; 如果都不成立, 则由公式 (21) 计算动态最优学习率

$$\eta^* = \frac{e^T \Phi \Phi^T e}{e^T \Phi \Phi^T \Phi \Phi^T e}$$

其中 $e = \hat{y} - y$, 更新连接权重 $W = W + \eta^* \Phi^T E$, $\Phi^T E = \Phi^T e$ 为成本函数的梯度;

步骤 4 更新网络输出值 $\hat{y} = \Phi W$, 更新实际输出值与网络输出值之间的均方根误差 $rmse$, 进入步骤 3;

步骤 5 输出最终的均方根误差 $rmse$ 、网络输出值 \hat{y} 并计算迭代次数.

与传统固定学习率的 RBF 神经网络相比, 具有最优学习率的 RBF 神经网络在每一步迭代中都加入了最优学习率的计算, 使得学习率在所有的迭代过程中都是合适的, 所以这种学习率被称为“动态最优学习率”. 上述学习算法在初始化阶段省去了人为设置学习率的步骤, 因而在使用过程中会更为简便.

2 实 验

在这一部分, 将进行金融时间序列预测实验来验证具有动态最优学习率的 RBF 神经网络的预测效果. 近年来金融时间序列分析引起了我国学者的广泛研究, 周炜星^[21] 利用拓扑学方法证明上证指数高频数据不具有多重分形的特性, 王春峰等^[22] 使用了小波分析的方法识别我国股市的价格跳跃. 作为一种智能方法, RBF 神经网络在金融时间序列分析领域也有着较为广泛的应用^[9], 但由于传统固定学习率的 RBF 神经网络不但要面临学习率的选择问题, 而且收敛速度较慢, 所以无法满足实际问题的实效要求, 在应用中受到了很大的制约. 利用具有动态最优学习率的 RBF 神经网络对沪深 300 指数日度波动率进行预测, 同时针对具有最优学习率的 RBF 神经网络的运行状态及预测结果与固定学习率的 RBF 神经网络以及其他两种神经网络(基于添加隐含层节点的 RBF 神经网络、前向 BP 神经网络) 进行对比. 所使用计算机的基本配置为: Intel® Atom™ CPU N280, 主频 1.67GHz, 内存 1GB; 所使用的模拟软件为 Matlab7.0.4.

实验所使用的数据为 2005 年 1 月 4 日至 2010 年 9 月 30 日沪深 300 日度波动率数据, 其中 2005 年 1 月 4 日至 2008 年 12 月 31 日的数据用于网络

训练与学习 2009 年 1 月 5 日至 2010 年 9 月 30 日的数据用于验证网络的预测效果. 为了简化算法和便于对比, 网络的具体参数设置如下:

$rmse^*$ —— 学习精度 (均方根误差) 设为 0.01;

c —— 隐含层节点中心用正交最小二乘法选定, 容差^[10] 为 0.1;

W_0 —— 隐含层到输出层的连接初始权重全部设为 0;

$\max t$ —— 最大迭代次数设为 50 000.

在网络学习阶段, 自变量为当期波动率的前 1 - 4 期数据生成的输入矩阵, 输出为当期波动率 (即使用前 1 - 4 期数据预测当期数据). 网络运行情况及其预测结果如表 1 所示.

表 1 沪深 300 指数日度波动率预测结果对比

Table 1 Results comparisons of HuShen 300 index daily volatility forecasting

	学习目标	学习率	运行时间	迭代次数	预测精度
具有最优学习率的 RBF 神经网络	0.01	动态最优	s^*	89*	0.020 2*
具有固定学习率的 RBF 神经网络 (学习率: 0.02)	0.01	0.02	INF	INF	INF
具有固定学习率的 RBF 神经网络 (学习率: 0.001)	0.01	0.001	62.434 041 s	1993	0.020 2
具有固定学习率的 RBF 神经网络 (学习率: 0.002)	0.01	0.002	50.428 895 s	997	0.020 2
基于添加隐含层节点的 RBF 神经网络	0.01	Null	123.570 084 s	—	0.021 9
前向 BP 神经网络	0.01	0.01	60.841 414 s	—	0.024 7

注: 基于添加隐含层节点的 RBF 神经网络及前向 BP 神经网络的迭代方式有所不同, 故在表中略去迭代次数

从实验结果中可以看出, 与传统固定学习率的 RBF 神经网络相比, 具有最优学习率的 RBF 神经网络并没有提高预测精度, 这是因为不论最优学习率还是固定学习率的 RBF 神经网络, 其初始参数都是相同的 (包括隐含层节点、初始权重和网络半径), 因而两者在达到学习目标后的网络形式是相同的, 所以造成了预测精度相同.

但同时可以看到, 具有最优学习率的 RBF 神经网络迭代次数明显减少, 运行效率有所提高, 同时也

保证了网络学习的稳定性. 可能的原因分析如下:

1) 相比固定学习率而言, 动态最优学习率能够保证每一步迭代都有合适的学习率. 截取了具有动态最优学习率的 RBF 神经网络学习过程的前 18 次迭代学习率数据 (见表 2), 从表 2 中可以看出, 每次迭代的学习率均不相同, 最大值为 0.097 7, 最小值为 0.010 38. 这样的学习率保证了网络能够以最少的迭代次数达到收敛, 同时避免了学习率过大而导致网络无法学习的问题.

表 2 前 18 次迭代的动态最优学习率

Table 2 The dynamic optimum learning rates in the first 18 iterations

迭代次数	1	2	3	4	5	6
学习率	0.097 7	0.008 77	0.051 616	0.009 735 1	0.034 675	0.010 283
迭代次数	7	8	9	10	11	12
学习率	0.032 252	0.010 367	0.032 02	0.010 377	0.032 013	0.010 38
迭代次数	13	14	15	16	17	18
学习率	0.032 026	0.010 382	0.032 04	0.010 383	0.032 054	0.010 384

2) 动态最优学习率在每一次迭代时都需要重新计算, 比固定学习率多消耗一部分计算时间, 但具有最优学习率的 RBF 神经网络在迭代次数

上已有相当大的缩减, 而最优学习率的计算又相对简单, 所以从总体上看, 具有最优学习率的 RBF 神经网络能够获得非常短的运行时间, 提高网络

的运行效率.

同时,使用固定学习率的 RBF 神经网络需要对不同的学习率进行不断尝试,以找到相对较优的学习率.这样的过程会耗费大量的时间,而具有最优学习率的 RBF 神经网络则简化了这一过程,更加便于使用.

对于基于添加隐含层节点的 RBF 神经网络,其每一次迭代都添加一个隐含层节点,权重则由公式(9)求逆获得,即

$$W = \Phi^{-1}y$$

当网络的输入数量较多时,为了达到学习目标,网络在学习过程中不得不添加大量的隐含层节点,这使得权重的求取变成了庞大的矩阵计算,严重影响了网络的运行效率.

对于前向 BP 神经网络,隐含层节点的数量需要自己设定.设置的过小,网络学习速度得以保证,但学习目标难以达到,预测精度也会受到影响;设置的过大,网络学习又会非常缓慢.在输入数量较多时,前向 BP 神经网络的学习速度和预测精度难以两全.

从表 1 中的预测结果可以看出,与以上两种神经网络及固定学习率的 RBF 神经网络相比,具有最优学习率的 RBF 神经网络不但能够保证较高的预测精度,同时也能节省大量的网络学习时间.

使用已学习好的具有最优学习率的 RBF 神经网络对 2010 年 10 月 8 日至 2010 年 12 月 31 日的沪深 300 指数日度波动率进行预测(结果见图 2)图中虚线部分为预测数据.从图中可以观察到网络对未来沪深 300 指数波动率的预测情况.

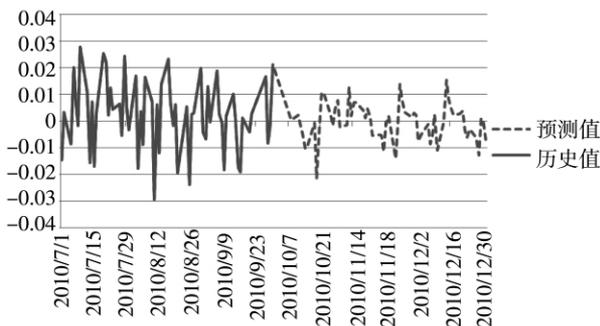


图 2 具有最优学习率的 RBF 神经网络对未来沪深 300 指数波动率的预测结果

Fig. 2 HuShen 300 index daily volatility forecasting using RBFNN with dynamic optimum learning rates

3 结束语

利用梯度下降法及简单的优化方法推导出了 RBF 神经网络的动态最优学习率,并将其应用于金融时间序列预测,在原有 RBF 神经网络的学习算法迭代过程中加入了求取最优学习率的步骤.因为动态最优学习率对于每一步迭代来说都是合适的,因此可以避免迭代次数过多或者是网络不学习等问题.

利用具有动态最优学习率的 RBF 神经网络学习算法,通过沪深 300 指数波动率预测实验可以发现:

(1) 使用动态最优学习率的 RBF 神经网络进行金融时间序列预测可以降低计算过程中的迭代次数,节约计算时间,同时可以保证网络正常学习.

(2) 使用固定学习率的 RBF 神经网络则面临学习率的选择问题,学习率过小会导致冗长的迭代,而学习率过大又会致使网络无法学习.

(3) 使用动态最优学习率的 RBF 神经网络可以达到比前向 BP 神经网络及基于添加隐含层节点的 RBF 神经网络更小的均方根误差,运行时间也更短.因此,动态最优学习率的 RBF 神经网络在实际应用中有着更为突出的优势.

动态最优学习率的推出,不但可以完善和改进 RBF 神经网络在金融时间序列预测方面的应用,还可以应用在其他方面,如工程预测、系统辨识等等.举例来说,在工程预测方面,动态最优学习率的 RBF 神经网络可以更加精确地预测出工程元件的劣化过程,可以为实际工程建设节省大量的成本.在系统辨识方面,RBF 神经网络本身就具有识别准确度高的特点,而具有动态最优学习率的 RBF 神经网络则能够更加迅速、准确地进行识别,将在很大程度上提高系统辨识的工作效率.

对于具有最优学习率的 RBF 神经网络学习算法来说,未来还可以做很多工作对其进行完善,比如如何确定最优的隐含层节点数,怎样将精度进一步提高,这些都可以成为今后研究的方向.

参考文献:

- [1] Bouchoux S, Brost V. Implementation of pattern recognition algorithm based on RBF neural network [C]// *Advanced Signal Processing Algorithms, Architectures, and Implementations XII*, 2002, 4791: 125 – 135, 468.
- [2] Zhao Z Q, Huang D S, Sun B Y. Human face recognition based on multi-features using neural networks committee [J]. *Pattern Recognition Letters*, 2004, 25(12): 1351 – 1358.
- [3] Haddadnia J, Faez K, Ahmadi M. A fuzzy hybrid learning algorithm for radial basis function neural network with application in human face recognition [J]. *Pattern Recognition*, 2003, 36(5): 1187 – 1202.
- [4] Li M, Tian J, Chen F. Improving multiclass pattern recognition with a co-evolutionary RBFNN [J]. *Pattern Recognition Letters*, 2008, 29(4): 392 – 406.
- [5] Lucks M B, Oki N. A radial basis function network (RBFN) for function approximation [C]. *42nd Midwest Symposium on Circuits and Systems*, 1999, 1 – 2: 1099 – 1101, 1150.
- [6] Lu Y, Sundararajan N, Saratchandran P. A sequential learning scheme for function approximation using minimal radial basis function neural networks [J]. *Neural Computation*, 1997, 9(2): 461 – 478.
- [7] Park J, Sandberg I W. Universal approximation using radial-basis-function networks [J]. *Neural Computation*, 1991, 3(2): 246 – 257.
- [8] Park J, Sandberg I W. Approximation and radial-basis-function networks [J]. *Neural Computation*, 1993, 5(2): 305 – 316.
- [9] Yu L, Lai K K. Multistage RBF neural network ensemble learning for exchange rates forecasting [J]. *Neurocomputing*, 2008, 71(16 – 18): 3295 – 3302.
- [10] Chen S, Cowan C F N, Grant P M. Orthogonal least squares learning algorithm for radial basis function networks [J]. *IEEE Transactions on Neural Networks*, 1991, 2(2): 302 – 309.
- [11] Wang M, Yang S, Wu S, et al. A RBFNN approach for DoA estimation of ultra wideband antenna array [J]. *Neurocomputing*, 2008, 71(4 – 6): 631 – 640.
- [12] Pei J S, Mai E C, Wright J P, et al. Neural network initialization with prototypes function approximation in engineering mechanics applications [C]// *2007 IEEE International Joint Conference on Neural Networks*, 2007, 1 – 6: 2110 – 2116, 3128.
- [13] Orr M J L. Introduction to radial basis function networks [EB/OL]. <http://www.anc.ed.ac.uk/mjo/intro/intro.html>, 1996.
- [14] Lee D W, Lee J. A novel three-phase algorithm for RBF neural network center selection [J]. *Advances in Neural Networks*, 2004, 3173: 350 – 355.
- [15] Husain H, Khalid M, Yusof R. Nonlinear function approximation using radial basis function neural networks [C]. *2002 Student Conference on Research and Development*, 2002, Proceedings: 326 – 329, 522.
- [16] Snyman J A. *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms* [M]. New York: Springer, 2005.
- [17] Sha D, Bajic V B. An online hybrid learning algorithm for multilayer perceptron in identification problems [J]. *Computers and Electrical Engineering*, 2002, 28: 587 – 598.
- [18] Yu L, Wang S Y, Lai K K. A novel adaptive learning algorithm for stock market prediction [J]. *Lecture Notes in Computer Science*, 2005, 3827: 443 – 452.
- [19] Sha D, Bajic V B. Adaptive on-line ANN learning algorithm and application to identification of non-linear systems [J]. *Informatica*, 1999, 23: 521 – 529.
- [20] Sha D, Bajic V B. On-line adaptive learning rate BP algorithm for MLP and application to an identification problem [J]. *Journal of Applied Computer Science*, 1999, 7(2): 67 – 82.
- [21] 周炜星. 上证指数高频数据的多重分形错觉 [J]. *管理科学学报*, 2010, 13(3): 81 – 86.
Zhou Weixing. Illusionary multifractality in high-frequency data of Shanghai stock exchange composite index [J]. *Journal of Management Sciences in China*, 2010, 13(3): 81 – 86. (in Chinese)

- [22]王春峰,姚 宁,房振明. 基于小波变换的多尺度跳跃识别与波动性估计研究[J]. 管理科学学报,2010,13(10): 63-68.
Wang Chunfeng, Yao Ning, Fang Zhenming. Research of multi-scale jump identification and volatility estimation based on wavelet transform[J]. Journal of Management Sciences in China, 2010, 13(10): 63-68. (in Chinese)

A RBF neural network with optimum learning rates and its application

WEI Min², YU Le-an^{1,2}

1. School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China;
2. Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100190, China

Abstract: In this study, a radial basis function (RBF) neural network learning algorithm with optimum learning rate is proposed. In this learning algorithm, the dynamic optimum learning rates which are determined by gradient descent and classical optimization technique are used to adjust the weight changes of RBF neural networks in an adaptive way. Using the dynamic optimum learning rates, the RBF neural networks can learn faster and more stable than the RBF neural networks with fixed learning rates. In order to verify the effectiveness of the proposed algorithm, volatility forecasting experiments on HuShen 300 index in Chinese stock market are conducted. The experimental results show the proposed RBF neural network learning algorithm with dynamic optimum learning rates could learn faster and avoid subjective selection of learning rates.

Key words: RBF neural networks; dynamic optimum learning rate; gradient descent

(上接第39页)

- [16]Belle G. Statistical Rules of Thumb[M]. New Jersey: John Wiley & Sons, 2008: 22-52.

Time-varying information spillover tests and their application to financial markets

LU Feng-bin¹, HONG Yong-miao²

1. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China;
2. Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen 361005, China

Abstract: This paper proposes two types of time-varying information spillover tests by combining Haugh and Hong statistics with rolling window method, and gives the selection rules of the rolling window size. Monte Carlo simulations show that both tests have good performances, and Hong's time-varying tests show better performance. The empirical study on the time-varying information spillover between Shanghai Futures Market (SHFE) and London Metal Exchange (LME) copper futures market show that the information spillovers between the two markets are apparently time-varying and SHFE's role in global copper markets raises gradually.

Key words: Granger test; information spillover; time-varying character; rolling window method; copper futures markets