

# 改进的 RBF 神经网络算法在金融时序序列预测中的应用<sup>①</sup>

## Application of Adjusted RBF Neural Networks to Financial Time Series Data Forecasting

刘淑梅 朱一嘉 许南山 (北京化工大学 信息科学与技术学院 北京 100029)

**摘要:** 在混沌理论的基础上,利用改进的 RBF 神经网络算法对金融时序序列数据进行预测,训练过程中采用动态方法调整径向基函数的中心点和宽度。最后通过对股票数据进行多日滚动预测证明该算法较改进前提高了预测的准确度,缩短了训练时间。

**关键字:** 径向基函数 最近邻聚类 最小二乘法 时序序列 混沌理论

### 1 引言

金融市场是一个非常复杂的非线性系统,经常受到政治,经济等多方面的影响。采用传统的预测方法对金融时序序列数据进行预测往往效果不好,文献[1]中提到用训练好的神经网络预测未来的混沌序列值有很好的预测效果。文献[2]认为金融时序序列数据是混沌的,是一个确定性的非线性系统,因此只要找到它内部的函数就可以对未来的值进行预测,由于金融市场的复杂性,我们很难找到这个函数,但是如果我们可以逼近这个函数,那么对金融时序序列数据进行短期的预测就是可行的。神经网络由于其自身优良的自学习,自适应能力以及良好的容错能力,可以逼近任意阶的非线性函数,这使其在金融时序序列预测中发挥着重要的作用。

RBF 神经网络具有学习收敛速度快、不会陷入局部最小值等优点,在时序序列分析预测、系统建模、故障诊断等领域广泛应用<sup>[3]</sup>。本文将改进的 RBF 神经网络用于金融时序序列预测,实验结果显示该算法在金融时序序列预测中有较好的预测效果。

### 2 预测的思路

① 单步预测: 设定历史关联天数为  $n$ , 即以前  $n$  天历史数据  $X_1, X_2, \dots, X_n$  作为输入, 预测第  $n+1$  天的输出。

从数学上讲即存在一函数  $P(t+1)=f(p(t), p(t-1)\dots p(1))$ 。但此方法对于金融时序序列预测来说实际意义不大。

② 滚动预测: 先进行单步预测, 再把输出反馈到输入端作为下一步预测输入的一部分, 以此来预测未来几天的价格走势。此方法适用于金融时序序列预测, 精度较高。

### 3 RBF网络的学习过程

#### 3.1 RBF网络结构

RBF 网络是一个三层的前向网络, 其结构如图 1 所示(假定只有一个输出), 从左到右分别为输入层, 隐含层和输出层。输入层把输入变量传到隐含层, 隐含层的基函数一般采用高斯函数, 对输入变量产生局部的响应, 隐含层到输出层是一个线性加权的过程。径向基函数为:

$$G(X_k) = \exp\left(-\frac{\|X_k - C_i\|^2}{r^2}\right) \quad (1)$$

RBF 网络输出为:

$$y_k = \sum_{i=1}^m w_i G_i(X_k) \quad (2)$$

① 收稿时间:2009-03-03

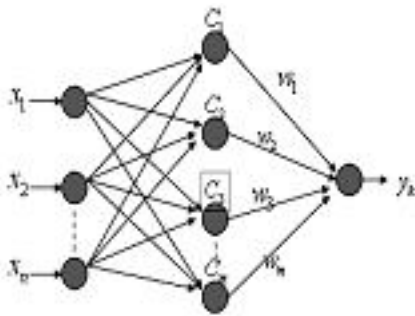


图1 RBF网络模型

## 4 算法的改进

针对上述提到的问题,对算法做如下的改进:

① 对径向基宽度  $r$  采用迭代的方法,设定增量  $t$ ,  $r$  的最小值  $\min r$  和最大值  $\max r$  ( $\min r$  和  $\max r$  的选取可见文献[6])。令  $r = r + t$ ,  $h$  表示迭代次数。对每次迭代的  $r$  值用改进的最近邻聚类算法对所有的输入数据进行分类,确定相应  $r$  值对应的中心点和权值。然后对每一个  $r$  值和相应的中心点,权值集合用训练数据计算模型的误差平方和:  $E_i = \frac{1}{n} \sum_{k=1}^n (Y_k - y_k)^2$ ,

$i=0,1,2 \dots h-1$ 。其中  $Y_k$  为期望输出,  $y_k$  模型实际输出。设  $E_p$  是所有  $E$  中值最小的一个。则  $r = \min r + p \cdot t$ ,  $r$  值对应的中心点和权值确定为 RBF 网络的中心点和权值。

② 每个输入向量确定属于某个聚类后,将该聚类的中心点的值修改为属于该类的所有输入向量的平均值。每输入一个向量就动态调整该类的中心点向量。

③ RBF 网络的中心点和径向基宽度确定后就不再修改了,而对连接权值  $w$  用 LMS 算法进行调整。采用批处理学习方法,把所有训练数据输入后计算平均误差平方和:

$$E(n) = \frac{1}{N} \sum_{i=1}^N (Y_i(n) - y_i(n))^2 \quad (3)$$

$n$  表示第  $n$  次迭代

对隐含层与输出层之间连接权值做调整:

$$w_i(n+1) = w_i(n) + \eta * \frac{\partial E(n)}{\partial w_i} \quad (4)$$

$\eta$  表示学习速度,取  $0 < \eta \leq 1$ ; 当  $E$  小于指定的阈值  $\varepsilon$  即  $E < \varepsilon$  则停止。在本文令  $\varepsilon = 0.0001$ 。

## 5 实验与分析

### ① 数据来源

本实验所有数据均来自 <http://www.wstock.net/wstock/5min.htm>。选取了中国石油化工从 2005-1-3 到 2006-6-21、熊猫电子从 2006-1-3 到 2008-8-8、江西铜业从 2003-12-1 到 2006-2-13 这三支股票的收盘价为数据进行测试。

### ② 实验流程

第 1 步: 导入数据并进行归一化处理

RBF 网络的学习分两步进行: 第一步首先确定径向基函数的中心点  $C$  和宽度; 第二步采用最小二乘法对连接权值进行修改。径向基中心点和宽度的确定是整个网络性能好坏的关键,中心点选取不当构造出来的网络性能往往不好,不好的中心点可能带来数据上的病态。

### 3.2 算法描述

文献[4]中提到了四种确定中心点的算法。最近邻聚类算法(nearest neighbor clustering algorithm)是一种在线的自适应学习算法,不需要提前确定隐含层的节点数,完成聚类后所得到的网络是最优的,很适合应用于金融时序序列。详细算法可参考文献[4],其大致过程[5]如下: 设  $r$  已给定,对所有参与网络训练的数据对作如下的操作,将第 1 个数据对中的输入变量作为第一个隐含层的节点,输出变量作为连接权值; 从第 2 个数据起,依次计算该数据对中的输入变量与已有的各个中心点的距离,找出其中最小的距离,进行判断。如果这个距离小于给定的  $r$  值,则将该数据对归到最小距离的那个中心点所属的类中,对应的中心点的向量不变,连接权值更新为该类中所有数据对的输出向量的算术平均值。训练过程结束后确定了中心点和连接权值。

### 3.3 算法的缺陷

在上述算法中径向基宽度  $r$  需要事先给定,这对于复杂的金融系统给出最佳的值存在一定的难度,往往需要经验和主观的猜测。仅以第一个属于某类中心的输入向量做为该聚类的中心量也无法构建一个最佳的网络结构,产生的中心点并非是这个聚类的最优值,势必会对网络的性能造成影响。

第2步: 根据历史关联天数将数据分成输入输出数据对

第3步: 寻找最佳的径向基宽度和中心点, 确定初始的连接权值

第4步: 采用批处理学习的过程用 LMS 算法对连接权值  $w$  作调整

第5步: 作未来七天的滚动预测并画图

### ③ 预测结果与数据分析

在实验中以股票的收盘价作为训练数据(除去最后的七天数据), 最后七天的数据作为检验数据。在本次实验中设定历史关联天数为 7 天, 最大迭代次数为 3000 次, LMS 的学习步长设为 0.9。下面给出了对比图, 图中绿色的点线表示实际输出, 蓝色的点线表示算法改进后的滚动预测输出, 红色的点线表示算法改进前的滚动预测输出。纵轴表示股票值, 横轴表示天数。

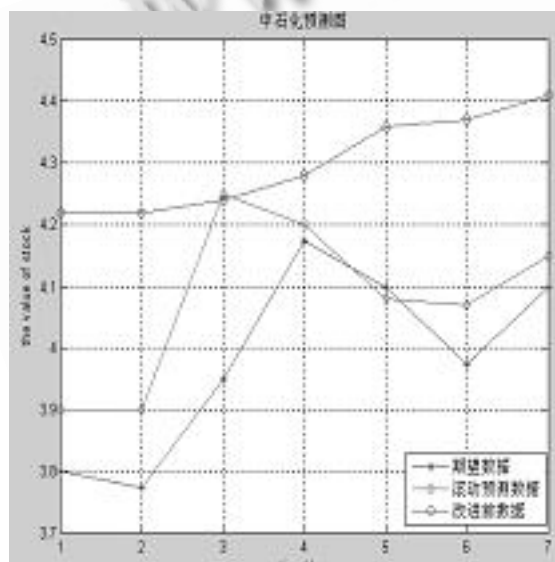


图 2 中石化预测曲线图

表 1 三支股票的预测精度和效率对比

| 股票   | 平均相对误差 |        | 训练用时 |        |
|------|--------|--------|------|--------|
|      | 改进后    | 改进前    | 改进后  | 改进前    |
| 中石化  | 1.48%  | 7.89%  | 1s   | 90.76s |
| 熊猫电子 | 0.72%  | 2.64%  | 2.6s | 19s    |
| 江西铜业 | 2.5%   | 22.67% | 3.5s | 316.5s |

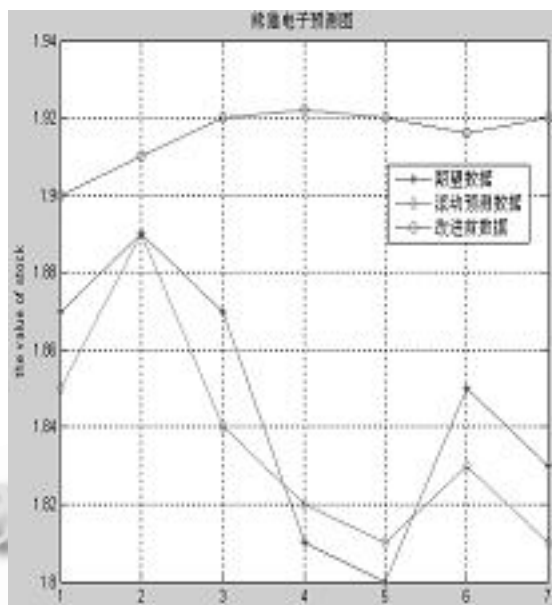


图 3 熊猫电子预测曲线图

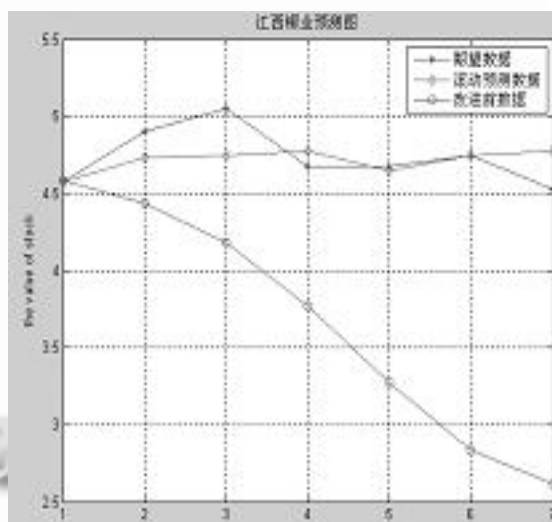


图 4 江西铜业预测曲线图

从上面三幅图中可以看出, 算法改进后预测的数据与实际值误差相差比较小, 预测数据的走势也和实际走势相吻合, 反观采用改进前的算法预测得出的数据与实际值的误差相对大一些, 并且预测的走势也与实际值不太吻合。

表 1 列出了采用改进后算法与改进前算法进行滚动预测的预测数据与实际值的平均误差和训练用时, 从数据对比中可以看出改进后的算法在精度和效率上

(下转第 186 页)

(上接第 178 页)

比原来的算法都有提高。可见用改进的 RBF 神经网络对金融时序列数据做滚动预测是切实可行的。

## 6 结论

本论文提出了一种基于 RBF 神经网络的金融时序列预测模型,通过对 RBF 学习算法的改进,提高了对金融时序列数据的预测准确度。选取最优径向基宽度和中心点的过程采用了迭代和动态更新的办法,因此需要系统记录每一次径向基宽度迭代时对应的中心点向量和输出层连接权值。本模型不仅可以对收盘价做预测,也可以对上证指数,开盘价等等其他的金融时序列进行预测。本文只考虑了收盘价这个因素,并没有对开盘价,最高价,最低价等相关数据进行分析。下一步的任务是把影响金融时序列的相关因素考虑进来,作为输入变量的一部分再进行预测。

## 参考文献

- 1 刘瑞平,沈福民.混沌时间序列预测与目标检测.雷达科学与技术, 2005,3(6):327-331.
- 2 杨一文,刘贵忠.基于神经网络的多变量时间序列预测及其在股市中的应用.信息与控制, 2001,30(5):413-417.
- 3 高隽.神经网络原理及仿真实例.北京:机械工业出版社, 2003:55-63.
- 4 朱明星,张德龙.RBF 网络基函数中心点选取算法的研究.安徽大学学报(自然科学版),2000,24(1):72-78.
- 5 孙延凤,梁艳春,孟庆福.改进的神经网络最近邻聚类学习算法及其应用.吉林大学学报(信息科学版), 2002,20(1):63-66.
- 6 储岳中.改进的 RBF 神经网络在非线性系统中的应用.计算机技术与发展, 2008,18(3):196-199.