

## 基于 BP 神经网络模型的 铁矿石价格预测研究

兴证期货·研发中心

2018 年 12 月 28 日 星期五

金融工程研究团队

### 内容提要

刘文波

从业资格编号：F0286569

投资咨询编号：Z0010856

段宁

从业资格编号：F3048895

孙石

从业资格编号：F3042665

吴耀宏

从业资格编号：F3035075

周英

从业资格编号：F3038963

联系人

吴耀宏

021-20370974

[wuyh@xzfutures.com](mailto:wuyh@xzfutures.com)

- 人工神经网络 (Artificial Neural Network) 是模仿人类神经网络中神经元互联结构以及网络在接收外部信息时表现出的“响应”机制，人工搭建的一种仿生数据处理模型。BP 神经网络 (Back Propagation) 是一种按误差逆向传播算法进行训练的多层前馈人工神经网络，是目前应用最为广泛的神经网络模型之一，它处理能力强、信息存储量大、容错性好、适应能力强。
- 主成分分析法 (Principal Component Analysis, PCA) 是一种将多维指标转化为维度较少的综合指标的数据处理方法。它的本质是线性正交变换，通过保留对方差贡献较多的子集来实现数据降维。
- 我们通过铁矿石产业链框架寻找影响其价格的因子，使用主成分分析法将原始数据降维去除冗余信息，基于 BP 神经网络模型建立对期货周度平均收盘价的预测模型。实证结果表明，神经网络模型的学习能力较强，训练效果满足预期，但外推泛化能力仍不够好，在绝对数值上的拟合度不好，但是在价格涨跌的方向性判断上正确率比较理想。动态建模的方法能够减小绝对数值拟合上的误差。

## 1. BP 神经网络模型

### 1.1 人工神经网络概述

人工神经网络 (Artificial Neural Network) 是模仿人类神经网络中神经元互联结构以及网络在接收外部信息时表现出的“响应”机制，人工搭建的一种仿生数据处理模型。通过神经网络的自我学习能力，它可以实现高性能的并行信息处理，对大量数据进行分析模拟。经过不断发展，神经网络已经在模式识别、自动化控制、分析信号等诸多领域有着重要的应用，基于神经网络的预测研究也在产量、物流、财务、风险预警等方面得到了广泛的实践。

对人工神经网络来说，节点和权值是它的核心要素，神经元节点通过设置的传递函数，对输入节点的数据进行相应的变换后再继续传递，权值则表示节点间连接强度的大小。确定了节点的连接方式和权值后便形成不同的拓补结构，将经过预处理的样本输入网络，网络会将对应样本的输出与目标真实值进行比较，然后通过调整连接权值，使输出接近预定的目标，这一过程就是神经网络的训练。训练完成即意味着一个网络拓补结构的建立，信息则以权值的形式存储在了这个网络中。因此，神经网络具有很强的可塑性，会随着外部环境的变化调整权值，从而改变映射关系训练成适应新变化的结构。

根据网络的结构和信号流的传递方向，神经网络可分为前馈网络和反馈网络。对于前馈网络，信号流从输入层单向地经过隐含层传递到输出层；而对于反馈网络，信号流会从一个神经元输出的方向再次传递回来，形成反馈回路。

### 1.2 BP 神经网络的定义和特点

BP 神经网络 (Back Propagation) 是 Rumelhart 和 McClland 等科学家于 1986 年提出的一种按误差逆向传播算法进行训练的多层前馈人工神经网络，是目前应用最为广泛的神经网络模型之一，它能够不通过事先确定函数关系式，只通过数据训练，将数据的输入及其对应输出之间的关系存储下来。BP 神经网络中主要包含两个过程，即信息的正向传播和误差的逆向传播。信息正向传播时，信息经由输入层传递到隐含层，在隐含层通过特定传递函数逐层对信息进行变换处理后，再传递到输出层。当输出值与期望值的差距不符合设定的训练目标时，误差信息将以某种形式经过隐含层向输入层逆向传播，并以此为依据逐层修正各单元的权值。信息不断传递和权值不断调整的过程就是神经网络的训练过程。当输出误差达到设定的目标范围内，或迭代次数达到了预设的学习次数，训练过程则会自动终止。

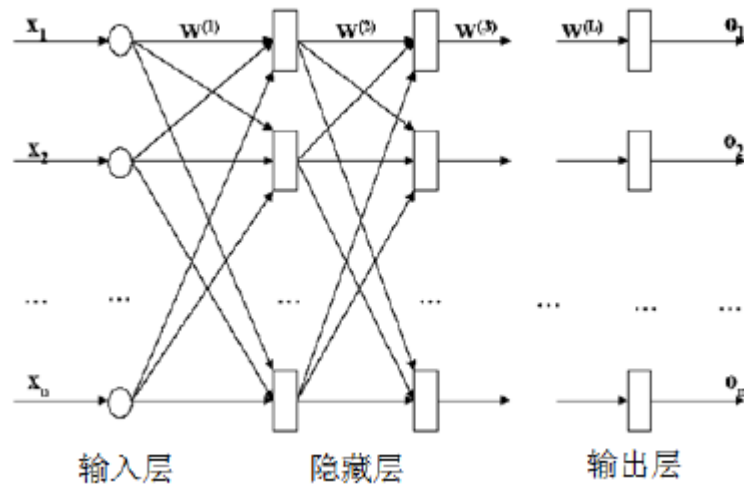
BP神经网络的数据处理方式的特点包括：

1. 处理能力强。BP神经网络通过给定样本进行自我学习，能够对信息并行处理，能逼近任意非线性函数关系，且泛化能力较好。
2. 信息存储量大。输入BP神经网络的信息以权值的形式存储在网络的神经元节点上，具有良好的记忆性。
3. 容错性好。BP神经网络模拟生物神经系统自动修复的特性，部分神经元的损伤不影响整体误差。
4. 适应能力强。BP神经网络通过其强大的自我学习和组织能力，能够不断调整各节点的权值来适应外部环境的变化。

### 1.3 BP神经网络的结构与原理

BP神经网络包含输入层、隐含层和输出层等三个层次。隐含层可以根据实际需要设置为单个或者多个隐含层。在节点之间的相互关系上，如图1所示，相邻层之间的节点，上一层的每一个节点都与下一层的所有节点相互连接，而同一层的节点之间则互不连接。

图 1: BP神经网络结构图

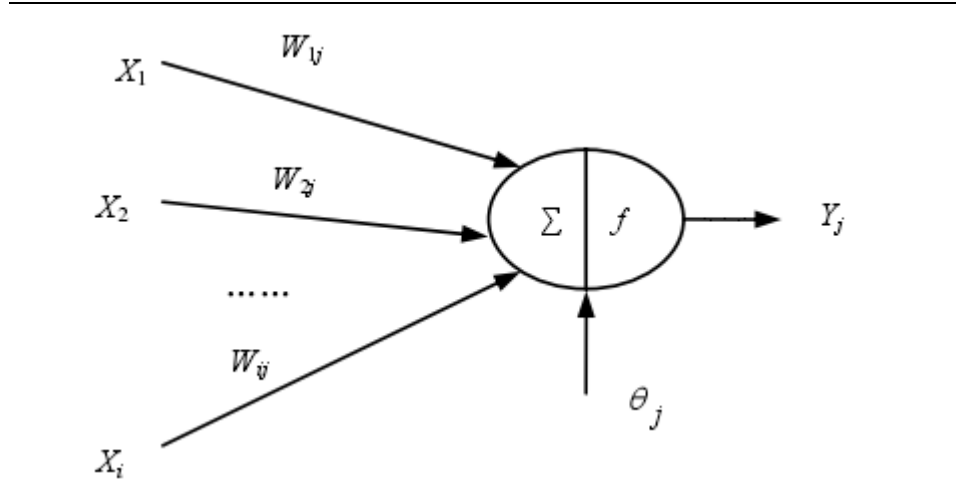


节点是神经网络的一个基本处理单位，承担着将来自上一层所有节点的信息输入转化为若干输出的功能，信息输入和输出的维度分别对应其上一层和下一层的节点数。关于在节点中进行的信息处理过程，首先是对输入加权求和，然后减去阈值，再根据传递函数进行相应的函数变换，从而得到相应的输出。节点的基本结构如图2所示：

来自上一层的多个信息 $X_i (i = 1, 2, \dots, n)$ 输入节点  $j$ ， $W_{ij}$ 是上层节点与节点  $j$  的连接权值， $\theta_j$ 是阈值， $f$ 是传递函数，节点中实现的变换如下：

$$Y_j = f\left(\sum_{i=1}^n W_{ij}X_i - \theta_j\right)$$

图 2: BP 神经网络节点基本结构



假设输出层有  $j$  个节点,  $Y_j$  为其输出,  $d_j$  为其期望输出, 则网络的误差为:

$$e = \frac{1}{2} \sum_{j=1}^n (d_j - Y_j)^2$$

根据反向传播的误差, 网络将调整权值, 这里引入梯度的概念, 即误差对权值的偏导数, 网络将沿着梯度的方向反向进行调整。

假设加权后的输入值为  $y_j$ , 则:

$$y_j = \sum_{i=1}^n W_{ij}X_i - \theta_j$$

那么梯度值可以转化为:

$$\frac{\partial e}{\partial W_{ij}} = \frac{\partial e}{\partial y_j} \frac{\partial y_j}{\partial W_{ij}}$$

其中:

$$\begin{aligned} \frac{\partial e}{\partial y_j} &= \frac{\partial \frac{1}{2} \sum_{j=1}^n (d_j - Y_j)^2}{\partial y_j} = \frac{\partial \frac{1}{2} \sum_{j=1}^n (d_j - Y_j)^2}{\partial Y_j} \frac{\partial Y_j}{\partial y_j} \\ &= -(d_j - Y_j) f'(y_j) \triangleq -\delta_j \end{aligned}$$

这里我们将绝对偏差量与传递函数导数的乘积定义为变量  $\delta_j$ , 代表局部梯度的概念。

$$\frac{\partial y_j}{\partial W_{ij}} = \frac{\partial \sum_{i=1}^n W_{ij} X_i - \theta_j}{\partial W_{ij}} = X_i$$

假设 $\eta$ 为学习率，则权值修正量为学习率与梯度的乘积，即：

$$\Delta W_{ij} = -\eta \frac{\partial e}{\partial W_{ij}} = \eta \delta_j X_i$$

从式中可以看出，权值需要进行的调整量，等于学习率 $\times$ 局部梯度 $\times$ 这一权值所连接的来自上一层的输入信号。

## 2. 主成分分析法

主成分分析法 (Principal Component Analysis, PCA) 是一种将多维指标转化为维度较少的综合指标的数据处理方法。它的本质是线性正交变换，先将原数据变换到一个正交的坐标系中，在这一坐标系中，变量按照方差贡献值顺次排列，通过保留对方差贡献较多的子集来实现数据降维。这一保留下来的综合变量集合就是主成分。

在预测模型体系中，输入模型用于预测的每一个因素的历史观察值序列都是一个列向量，代表了对预测目标影响程度的一种描述。如果这些因素之间存在线性相关，那么这一组描述预测目标影响程度的因子中就存在信息冗余，因此，通过主成分分析法进行降维，可以去除信息中的冗余成分，并提高神经网络的稳定性和训练速度。

进行降维时，首先需要对原数据的矩阵进行正交变换，从而用一个新的角度来描述对预测目标的影响程度。在这一新视角中，原有的因素的不同线性组合构成了新的因素并取代原有因素。新的因素并不具有具体的经济学含义，只是一个抽象的数值指标，且彼此线性无关。在这些相互独立的新因素中，我们通过剔除作用较小的因素来实现降维。

那么，应该如何进行正交变换呢？首先，我们假设模型输入  $n$  个因素，则构成矩阵  $X = (X_1, X_2, \dots, X_n)$ 。假设存在变换  $P$ ，将  $X$  变换为  $Y$ ，即  $Y = PX$ 。 $X$  和  $Y$  的协方差矩阵分别记为  $C$  和  $D$ ， $t$  为时间长度，则有：

$$C = \frac{1}{t-1} XX^T$$

$$D = \frac{1}{t-1} YY^T$$

$$D = PCP^T$$

因为我们希望实现的变换是使得新的因素之间彼此线性无关，所以因素之间的协方差为 0，那么矩阵  $D$  应该是一个对角阵，则  $P$  应该是矩阵  $C$  的特征向量组成的矩阵，矩阵  $D$  对角线上的元素为矩阵  $C$  的特征值。将特征值贡献大的子集作为主成分，对应地取出  $P$  的子集  $P'$ ，则  $Y = P'X$  便是我们实现降维的方法。

## 3. BP 神经网络模型的构建

### 3.1 网络初始化

模型的构建首先从网络的基础设定开始，包括几个方面：

#### (1) 设置隐含层层数和节点数

BP 神经网络可以包含多个隐含层，但理论上，单隐含层的网络通过适当增加隐含层的节点数就可以实现任何非线性连续函数的映射，只有在学习不连续函数时才需要多个隐含层，因此对于中小型网络来说，单隐含层能够满足需要。

隐含层的节点数对网络的学习能力有较大影响，如果节点数太少，网络的容量便不足以存储训练样本中所蕴藏的信息，理论上节点数越多，对函数的逼近能力越强。但节点过多不仅造成网络训练计算量的增加，还会将样本中的干扰和噪声存储进去，造成过拟合，导致泛化能力的降低。因此，我们将在实践的基础上在拟合效果与节点数之间作出权衡。

#### (2) 网络节点的传递函数

误差反向传播算法涉及梯度的计算，这里要求传递函数必须连续可微，因此这里使用的是 Log-Sigmoid 函数，它具有饱和和非线性放大系数的功能，适合于处理和逼近非线性的函数关系。它的表达式为：

$$f(x) = \frac{1}{1 + e^{-x}}$$

### 3.2 样本划分

样本内数据通常会划分为两组，分别是训练样本和验证样本。在网络训练的过程中，训练样本直接参与学习过程，是连接权值变化调整的依据。验证样本不直接参与权值的调整，它的主要作用是防止网络过拟合。在网络训练时，用训练样本每训练一次，都会将验证样本数据输入网络进行验证，得到验证输出误差。当这一误差连续几次验证都不下降甚至上升，次数达到了事先设定的步数时，我们便认为训练误差已经不再减小，训练效果已经是所能达到的最优情形，此时应停止训练。如果此时继续训练，网络可能陷入过学习而导致泛化能力不好。

### 3.3 确定训练算法与训练参数

BP 神经网络存在多种训练算法，不同的训练方法有着不同的收敛速度和拟合性能。这里采用的是 LM 算法。

Levenberg-Marquardt (LM) 算法是在 Gauss-Newton 算法的基础上改进得到的，它的优点是对于过参数化问题不敏感，能有效处理冗余参数，

减小网络陷入局部最优的概率，更重要的是这一算法的收敛速度最快，且拥有较小的均方误差。

确定了训练算法后，还需要对其他训练参数进行设置，具体如下表所示：

表 1：模型训练参数

参数名称	参数设置
性能函数类型	Mean Squared Error (mse)
训练目标	$MSE < 10^{-3}$
最大训练次数	1000
最大验证失败次数	6
动量因子初值	0.001
训练方法	Levenberg-Marquardt 法

## 4. PCA-BP 模型的训练与预测

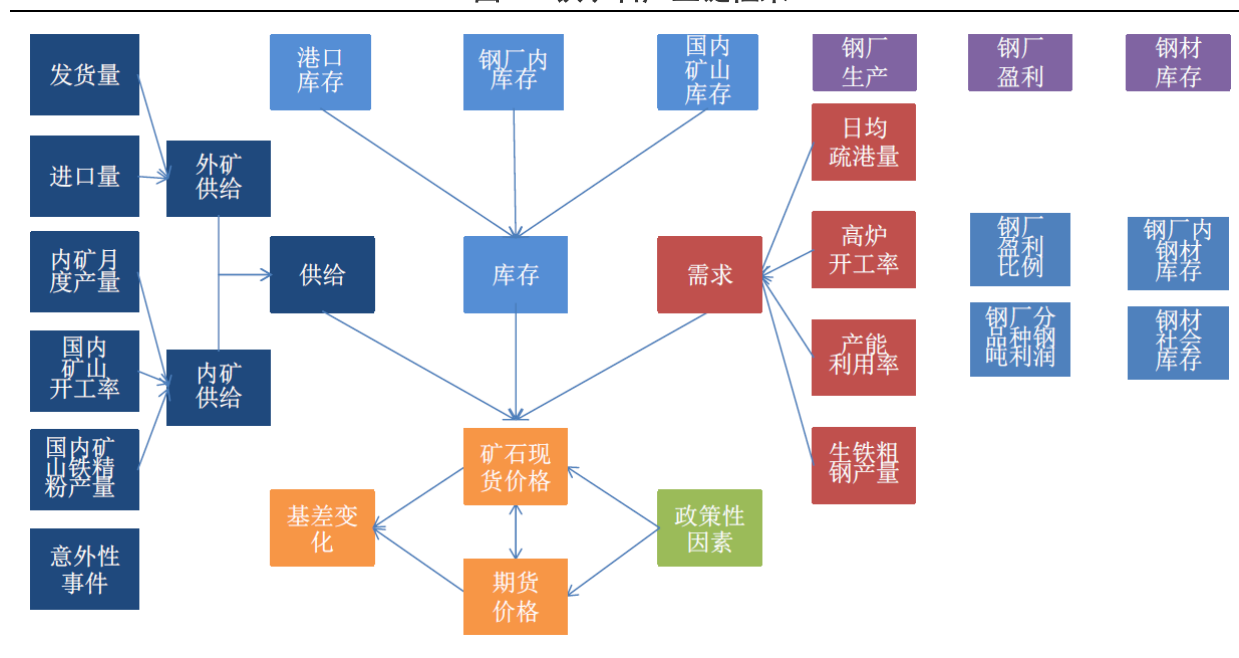
### 4.1 数据选取与预处理

铁矿石是指存在利用价值、含有铁元素或铁化合物的矿石。提炼 1 吨生铁约需要 1.6 吨铁矿石，在生铁成本中占比 50% 以上。

全球铁矿石分布集中，主要在澳大利亚、巴西、俄罗斯等国。我国铁矿石产量无法满足需求，因此有较大的进口依赖，铁矿石定价与海外矿山定价关系密切。由于铁矿石 98% 是作为生产钢铁的原材料使用，因此铁矿石的价格与下游钢铁行业有紧密关联。总的来说，铁矿石的价格主要受到来自上游矿山、海运以及下游钢铁行业等方面的影响。

如图 3 所示，铁矿石价格受到多方面的影响，考虑数据更新频率，我们选取了巴西澳洲发货总量、铁矿石港口库存、铁矿石日均疏港量、国内大中型钢厂进口铁矿石平均库存可用天数、矿山开工率、矿山铁精粉产量、矿山铁精粉库存、高炉开工率、钢厂盈利比率、粗钢预估日均产量、重点企业钢材库存、主要城市钢材库存、螺纹钢现货价格等指标，在此基础上再加上反映海运运费的波罗的海散干货指数这一指标，对铁矿石价格进行预测。因为数据的频率大部分是周度，因此我们也以周度为单位进行预测。预测目标使用铁矿石期货周度平均收盘价，并将其滞后一期的数据也作为因子之一。

图 3：铁矿石产业链框架



资料来源：兴证期货研发部

我们对数据的预处理主要包括：

(1) 因子频率调整

由于不同因子数据更新频率与具体时点各异，我们将其都统一成周度频率的数据，以期货平均收盘价序列为基础，将其他数据与之对齐，基本原则是数据无更新或缺失时以这之前最近一期的数据来填充。而对于日度数据我们则取其周度平均值形成一个新的周度序列。

(2) 期货价格连续化

我们用于计算周度平均的期货价格是一个主力合约连续化的价格，假设为主力合约切换日的收盘时切换到新主力合约，将主力合约价格序列转化为价格变化率序列后，再用价格变化率序列反推连续化的价格序列，从而对价格进行后复权。

(3) 样本归一化

由于不同指标的数量级差别较大，如果用原始数据直接预测，因子对预测目标的影响程度也差别较大，小数值信息有可能被大数值信息淹没。且由于传递函数饱和和非线性的特性，未经处理的数据会使得节点的输出落在饱和区域，使得网络学习速度慢而难以收敛。因此这里将样本数据归一化映射到[-1, 1]区间，从而使得样本不同维度的特征具有相同的尺度。归一化函数为：

$$y = \frac{2x - x_{max} - x_{min}}{x_{max} - x_{min}}$$



## 4.2 主成分降维

根据铁矿石产业链框架，我们的样本数据包括 16 个因子，将预处理后的周度数据组成  $n$  行 16 列的样本矩阵  $X$ ， $n$  为样本数：

$$X = \begin{bmatrix} X_{1(1)} & \cdots & X_{16(1)} \\ \vdots & \ddots & \vdots \\ X_{1(n)} & \cdots & X_{16(n)} \end{bmatrix}$$

求出样本的协方差矩阵：

$$C = \frac{1}{n-1} XX^T$$

接着求得协方差矩阵的特征值以及对应的特征向量，将特征值从大到小排列，因为特征值代表的是新因子的方差，因此我们要得到对总方差贡献率达到 95% 以上的特征值子集元素数量，即得到下式的  $i$ ：

$$\frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^{16} \lambda_k} \geq 95\%$$

将特征向量依照特征值的顺序排列，取前  $i$  列构成变换矩阵  $P$ ，则  $Y = PX$  就是降维后的样本数据矩阵。

根据程序运行结果，前 8 个特征值的方差贡献率之和达到 95% 以上，因此样本向量由 16 维降至 8 维。

## 4.3 PCA-BP 模型训练与测试结果

我们使用了从 2013 年 10 月到 2017 年 12 月的数据，经过预处理和主成分降维后，模型的输入节点为 8 个因素，输出节点为期货周度平均价格。

我们以 5 个节点作为起点，以 5 作为步长，在每个节点数上进行 1000 次训练，并记录每次训练得到的训练样本误差、总误差和网络输出值，将 1000 次的结果取平均来进行比较，结果如表 2 所示：

表 2：各节点数训练结果

节点数	训练样本误差	总误差	方向判断胜率
5	0.007698	0.006894	65.12%
10	0.004698	0.005154	66.98%
15	0.002109	0.00482	69.77%
20	0.001169	0.004819	72.09%
25	0.000877	0.004889	73.02%
30	0.000458	0.005267	74.42%

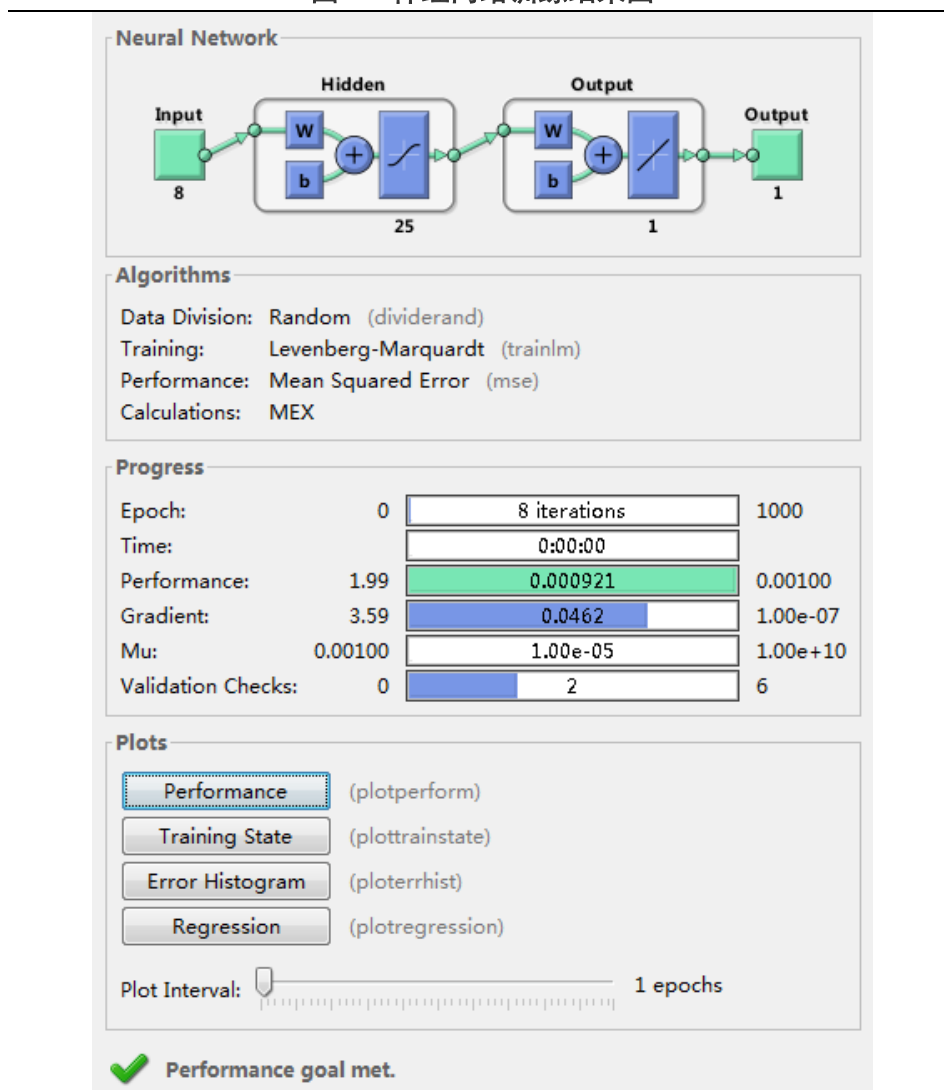
资料来源：兴证期货研发部

训练样本误差是训练样本拟合值与真实值之间的均方误差，总误差是

全样本网络输出与真实值之间的均方误差，这一误差来自于训练样本和验证样本。另外我们将全样本网络输出值逆归一化还原成期货价格，并与真实价格进行比较，统计其预测值的涨跌方向与真实值涨跌方向的一致情况。从表格数据可以看出，随着节点数的增加，网络的学习能力逐渐提高，训练样本的拟合度随之提升，表现为训练误差的降低和胜率提升，但总误差经历了一个下降再上升的过程，说明节点数过多会导致模型泛化能力下降。因此对于节点数的选择，我们在平均训练误差满足训练目标  $MSE < 10^{-3}$  的情况下，选择最小的节点数以保证一定的泛化能力，因此最终我们取节点数为 25。

下图为 25 节点时其中一次训练的结果图。从图中可以看到，网络仅经过八次迭代便收敛至训练目标，且验证次数只有 2，训练结果满足预期。

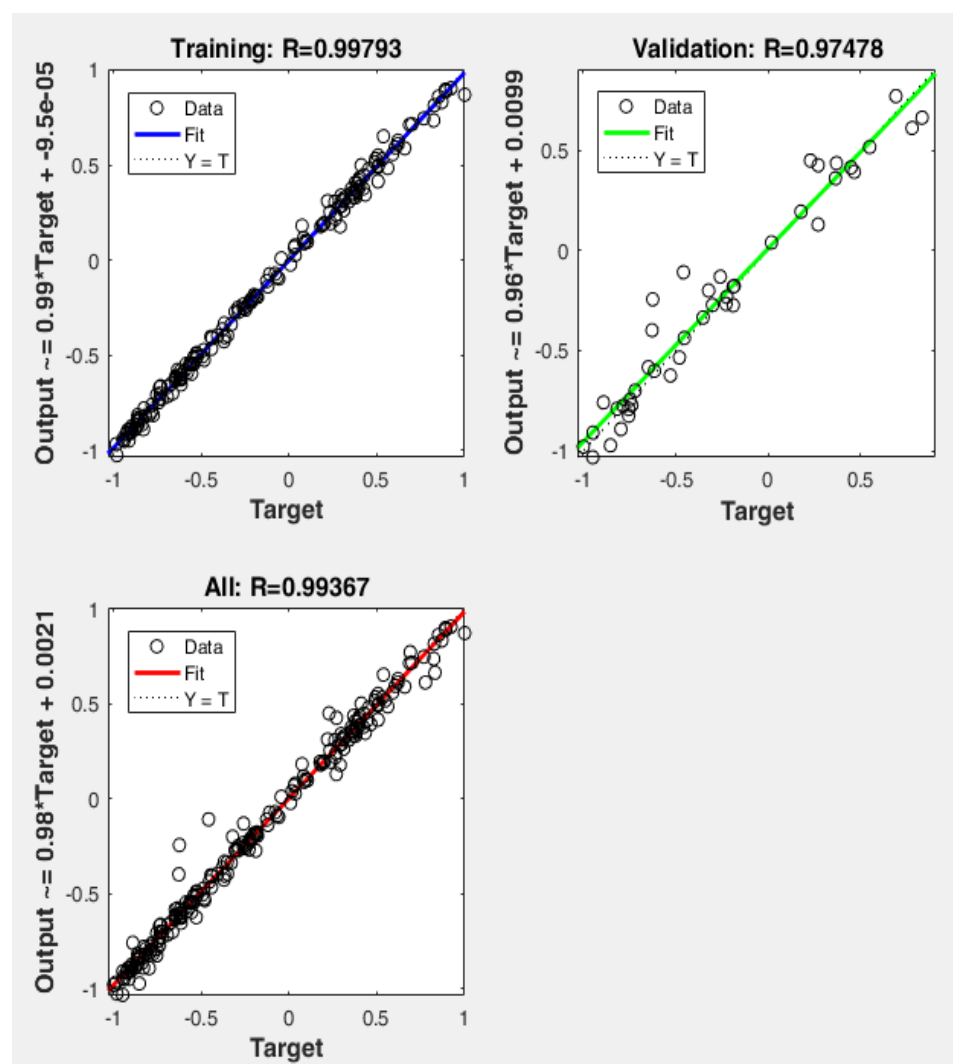
图 4：神经网络训练结果图



数据来源：兴证期货研发部

我们通过散点图来查看网络的拟合情况。从图 5 中可以看出，网络的总体拟合情况较好，训练样本拟合度为 99.79%，验证样本拟合度为 97.48%，全样本拟合度为 99.37%。

图 5：网络拟合情况



数据来源：兴证期货研发部

我们将 1000 次训练中样本对应网络输出的平均值与真实期货价格作对比，可以看到样本对应的网络输出与真实值的拟合程度较好，训练达到了我们的预期目标。

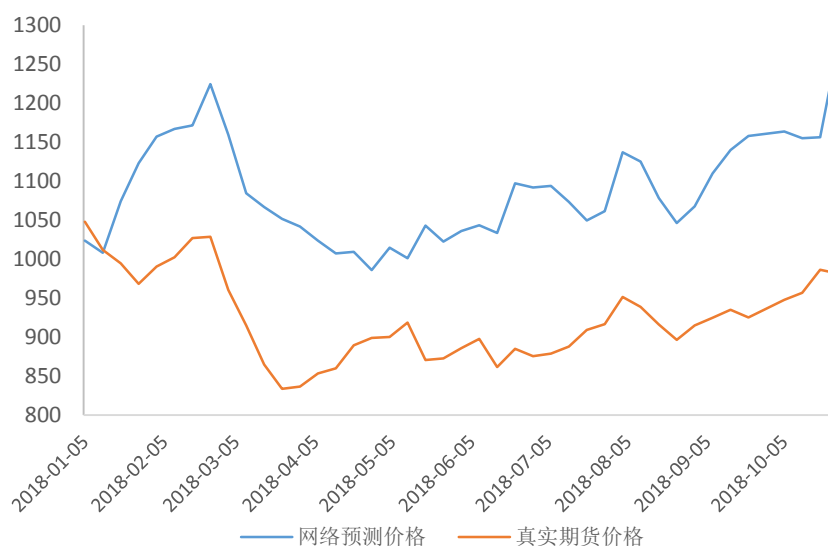
图 6：网络预测价格与真实价格的比较



数据来源：兴证期货研发部

接着我们取 2018 年 1 月到 11 月的数据作为测试样本，将测试数据输入训练好的网络中，得到网络输出值。这里我们同样进行 1000 次训练，并将测试样本输入每次训练得到的网络从而获得当次测试结果，将 1000 个测试结果取平均作为我们的最终结果。我们将测试样本对应的网络输出与真实值进行比较，结果如图 7 所示：

图 7：测试样本网络预测价格与真实价格的比较



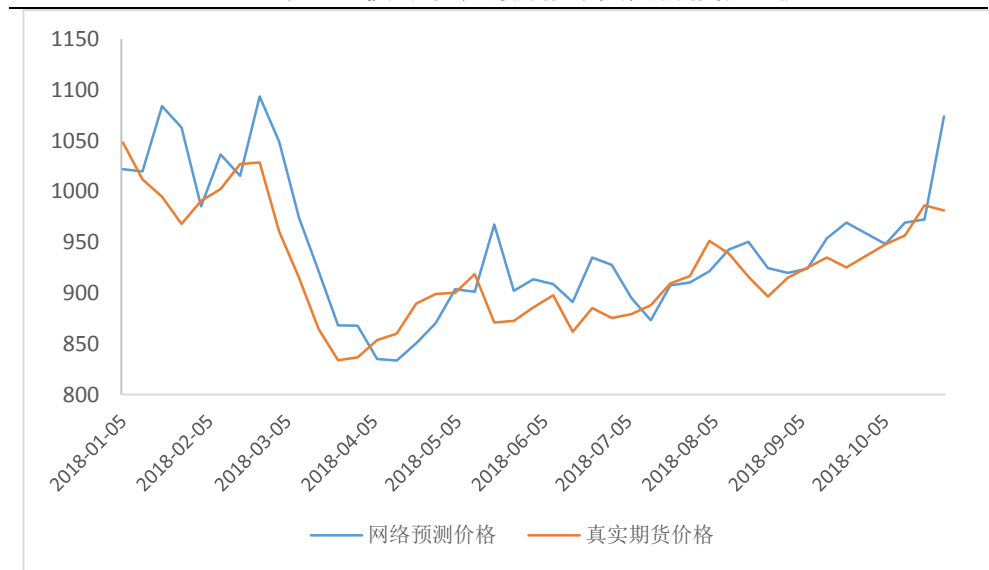
数据来源：兴证期货研发部

从图上来看，测试样本网络预测值在绝对数值上的拟合度并不好，但我们可以发现预测值的曲线走势与真实值一致性较高，因此我们对价格变动方向判断正确率进行统计，统计结果显示胜率是 65.85%。由于我们使用的期货价格是一个复权之后的序列，只保留了收益率关系，数值本身已经不是真实价格，因此方向性判断对于我们来说更有实际意义。

#### 4.4 动态建模

我们可以看到，模型在样本外推时，绝对数的预测上误差较大。因此我们尝试使用动态建模的方法使得网络能学习外界的最新变化。在动态建模法之下，训练的网络只用来预测下一期的数据，每预测新一期的数据时，都对包含了最新数据的样本重新进行归一化和主成分分析，再重新训练 1000 次并将产生的输出取平均作为我们的预测结果。我们同样从 2018 年 1 月开始进行预测，动态建模法下的预测结果与真实值对比如下图：

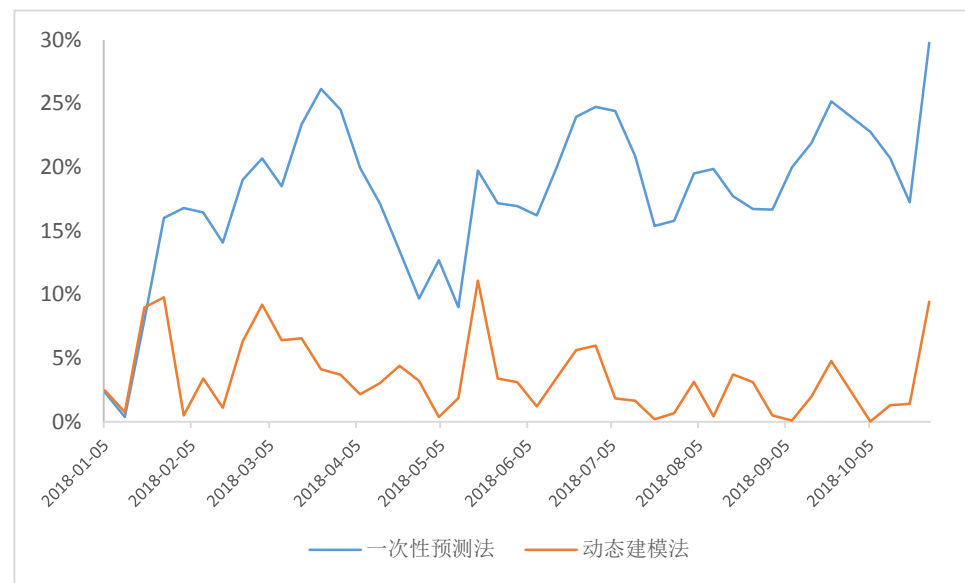
图 8：动态建模网络预测价格与真实价格的比较



数据来源：兴证期货研发部

可以看到动态建模下，网络输出值与真实值的拟合程度较理想。我们将两种方法下预测值与真实值的误差进行对比，可以看到，在动态建模下，绝对数值上的误差百分比有了明显的降低，平均误差在 3.5%。同样地，我们也对动态建模预测下价格变动方向判断的正确率进行统计，统计结果为 56.1%。

图 9：不同建模方法下预测值与真实值的绝对误差



数据来源：Wind，兴证期货研发部

## 5. 小结

本文通过铁矿石产业链框架寻找影响其价格的因子，使用主成分分析法将原始数据降维去除冗余信息，基于 BP 神经网络模型建立对期货周度平均收盘价的预测模型。实证结果表明，神经网络模型的学习能力较强，训练效果满足预期，但外推泛化能力仍不够好，在绝对数值上的拟合度不好，但是在价格涨跌的方向性判断上正确率比较理想。动态建模的方法能够减小绝对数值拟合上的误差。这一建模方法开启了一个新的思路，后续可推广到其他品种甚至其他预测标的上做更多尝试。

## 分析师承诺

本人以勤勉的职业态度，独立、客观地出具本报告。本报告清晰地反映了本人的研究观点。报告所采用的数据均来自公开资料，分析逻辑基于本人的职业理解，通过合理判断的得出结论，力求客观、公正，结论，不受任何第三方的授意影响。本人不曾因也将不会因本报告中的具体推荐意见或观点而直接或间接接收到任何形式的报酬。

## 免责声明

本报告的信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。文中的观点、结论和建议仅供参考。兴证期货可发出其它与本报告所载资料不一致及有不同结论的报告。本报告及该等报告反映编写分析员的不同设想、见解及分析方法。报告所载资料、意见及推测仅反映分析员于发出此报告日期当日的独立判断。

客户不应视本报告为作出投资决策的惟一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。

在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的损失负任何责任。

本报告的观点可能与资管团队的观点不同或对立，对于基于本报告全面或部分做出的交易、结果，不论盈利或亏损，兴证期货研究发展部不承担责任。

本报告版权仅为兴证期货有限公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用、刊发，需注明出处兴证期货研究发展部，且不得对本报告进行有悖原意的引用、删节和修改。